

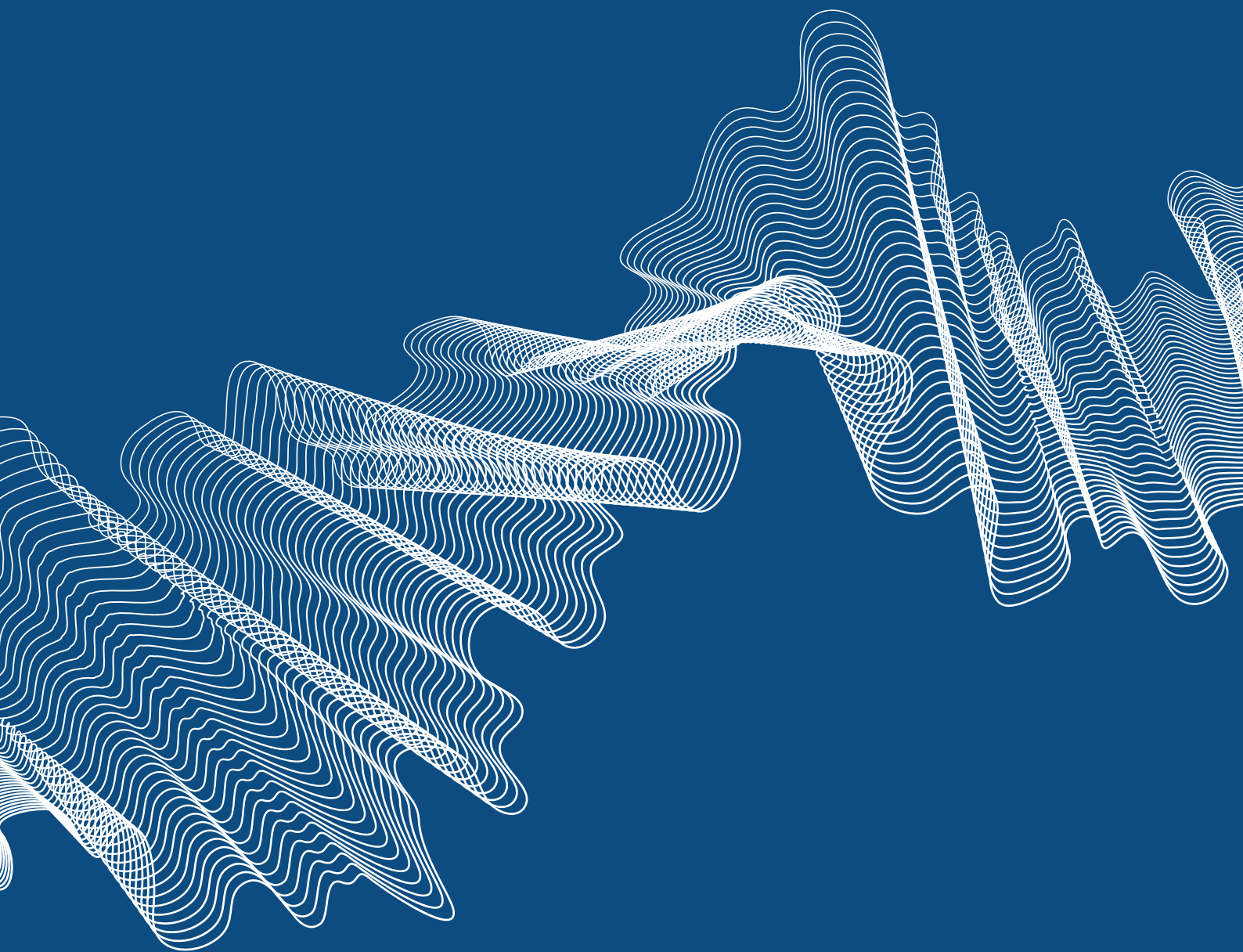
INFORMATION ECOSYSTEMS AND TROUBLED DEMOCRACY

A Global Synthesis
of the State of Knowledge
on News Media, AI
and Data Governance

*Robin Mansell, Flavia Durach,
Matthias Kettemann, Théophile Lenoir,
Rob Procter, Gyan Tripathi,
Emily Tucker*

JANUARY 2025

This report is dedicated to the
memory of Christophe Deloire,
late Chair of the Forum on
Information and Democracy



FOREWORD



**CAMILLE GRENIER, EXECUTIVE DIRECTOR
OF THE FORUM ON INFORMATION AND DEMOCRACY
AND IRIS BOYER, HEAD OF THE OBSERVATORY**

The current information and communication space is facing an unprecedented decline in public trust, marked by power asymmetries between tech companies and the rest of society, alongside a growing reliance on social media as a primary source of information. Information, communication and technology have never been so available at global scale, yet online citizens all over the world are experiencing an increasingly hostile environment impacted by information manipulation and weaponization, exploitative data practices, and the undermining of media pluralism and independence. Meanwhile, the rapid advance of disruptive frontier technologies such as AI risks outpacing ethical and regulatory safeguards.

These developments have tangible repercussions both for individuals and the institutions governing them. They are happening fast and globally. They are often met with under-evidenced reactive and siloed policy responses, or with technology-facilitated blanket solutions that fail to address specific regional and community needs.

To collectively and meaningfully address information chaos and democratic instability, a shared understanding is needed of how the evolving information ecosystem intersects with democracy.

The Partnership for Information and Democracy, endorsed by 55 countries, has established the principles that should govern a democratic information and communication space. The Observatory on Information and Democracy seeks to provide governments and the broader public with a clear understanding of the structure of this space and its effects on democracy. By gathering academics, activists, policy makers, regulators and tech representatives around this shared ambition, the Forum on Information and Democracy sought to build an IPCC equivalent for the information and communication landscape. By producing a regular global assessment on issues at the intersection of information and democracy, the Observatory strives to critically evaluate evidence from a range of disciplinary and regional perspectives, to establish the state of scientific knowledge, and to enable more meaningful cooperation between the academic and policy communities.

When we started this journey more than two years ago, we embarked on a thorough prefiguration process spearheaded by humbling thought leaders like Ángel Gurría, Shoshana Zuboff, Maria Ressa and the late Christophe Deloire. The ambition was huge and the resources scarce.

Unlike the IPCC for climate change, this initiative seeks to establish knowledge on emerging and highly debated issues investigated by a plurality of disciplines and perspectives, where evidence is scarce and data access limited. Despite material hardships, the collective mobilization of effort has been proportional to this generation-defining challenge and motivated us to be bold.

Inspired by the hundreds of written and oral contributions we received through calls for papers and the consultations we led with our Stakeholders Advisory Group and Steering Committee, we set a really high bar by planning to publish a global analysis of research covering a very wide array of relevant questions about **the Media, AI and Data Governance with a cross-cutting theme of mis- and disinformation.**

As a result of our collective engagement with three research assessment panels comprised of over 60 volunteer researchers, coordinated by six rapporteurs and led by a Scientific Director, this landmark report gathers **more than 1,600 sources.** The report tackles issues as diverse as: ***Trust in news and tech platforms' role in its evolution; how mis- and disinformation is linked to societal and political polarization; and how data governance can ensure justice in a data-led economy.***

This inaugural global assessment does not provide silver bullets to tackle these issues. It zeroes in on what we know and can agree on and what we don't know or don't agree about yet. It defines a clear pathway for future research and offers actionable insights for policymakers and tech company representatives. We hope that the report can be a scientific blueprint for multidisciplinary, collaborative and open research methodologies and that it can serve as a beacon for policy and action communities writ large.



COURTNEY RADSCH, CHAIR OF THE STEERING COMMITTEE

Director of the Center for Journalism and Liberty at the Open Markets Institute,
Fellow at UCLA Institute for Technology, Law and Policy and Fellow at the Center
for Democracy and Technology.

It has been a privilege to chair the Steering Committee of the Observatory on Information and Democracy, which provided guidance to this remarkable effort to survey the state of knowledge across various domains and from around the world to understand the complexity of information ecosystems and their profound implications for democratic societies. This report is not just an academic exercise—it is a roadmap for policymakers, researchers, and civic leaders committed to preserving and strengthening the democratic potential of our global information ecosystems.

Too often, policymakers grapple with isolated aspects of information systems without fully comprehending the broader contextual forces at play. Too much recent research has been overly fixated on social media and individual behaviors, overlooking the complex systems that shape how information is produced, distributed, and consumed and providing minimal guidance to policymakers who are rightly reticent to get involved in speech.

By adopting an «information ecosystem» framework, we provide a holistic analysis that reveals the profound interdependencies between people, institutions, technologies, and practices. This approach moves beyond simplistic explanations, offering policymakers a more nuanced understanding of how information systems actually function. Our report provides a rigorous, interdisciplinary assessment that transcends traditional silos and underscores the need for a wider array of policy interventions that proactively shape the public sphere rather than being held hostage to corporate power and reactive interpretations based on narrow perspectives. Our research illuminates how political economic structures, sociotechnical systems, and editorial practices are deeply interconnected, constraining policy options while simultaneously demanding more integrated and adaptive solutions. And it makes clear that the health of our information ecosystems cannot be reduced to combating mis- and disinformation through narrow, technology-centric interventions that fail to grapple with the wider societies. Perhaps most critically, our research reveals a troubling reality: the current state of information ecosystems is predominantly shaped by powerful corporate actors who achieve market dominance through strategic monopolization. This combination of regulatory complexity and corporate power has led to a form of “benign neglect” that undermines the integrity of our information environments. This narrow lens has significantly constrained our understanding of the multifaceted factors essential for cultivating healthy information systems that can sustain and strengthen democratic processes.

This report is an invitation to grapple with the complex interplay between infrastructure, media systems, civil society, and public sector institutions. Our findings suggest that effective policy solutions must:

- Acknowledge the historical, political, economic, and social forces that shape information flows and sociotechnical systems.
- Develop holistic approaches that consider the entire ecosystem rather than isolated interventions.
- Understand how data governance and AI systems fundamentally influence information production, dissemination, and consumption.

- Use a wider array of authorities and policy tools to create legal, regulatory and normative frameworks that protect democratic values while empowering communities and individuals who are embedded in increasingly opaque sociotechnical systems.

By exploring both individual-level phenomena and systemic dynamics, this report suggests how legislative, regulatory, competition, education and other public authorities as well as tech platforms and citizens themselves all have a role to play in cultivating information ecosystems where democracy can better thrive.

INSIGHTS FROM THE STEERING COMMITTEE



The report should be considered required reading for all policymakers engaged in efforts to implement commitments to a healthy information ecosystem. On the topic of AI, for instance, the report highlights the importance of being specific about the type and role of AI systems involved in the media landscape and the crucial role that lack of clear accountability over their governance plays in the failure to address persistent problems.

Ansgar Koene, Global AI Ethics and Regulatory Leader at EY and Senior Research Fellow at the Horizon Digital Economy Research institute (University of Nottingham).



From a scientific approach but with political will, the Observatory's inaugural report achieves several objectives: presenting this intricate landscape, separating the wheat from the chaff for clearer understanding, delving deeper to avoid superficial conclusions, and, no less importantly, proposing actionable steps based on this analysis.

Natalia Zuazo, Author of *Guerras de internet* and *Los dueños de Internet*, Director of SALTO Agency and UNESCO consultant.



It's been more than a decade since I enrolled into a university to study Digital Democracy. Between then and now, the influence of information systems on democracy has expanded beyond the imaginable. Trust in systems, digital, informational, governance and politics are challenged at all levels. The inaugural report of the International Observatory on Information and Democracy steps in at a critical time in human history when the emergence and evolution of artificial intelligence is challenging all human fundamentals of trust. In-depth research has gone into this report to bring to the fore what citizens, governments, and big tech can focus on when navigating the unsure waters of information systems in democracy.

The recommendations of the report help our orientation, especially as we grapple with the tasks of connecting the unconnected, sustaining traditional media while also being open to new media, striving to maintain digital inclusion while also incentivizing markets, upholding security while protecting human rights, and over all, striving to build trust across the board.

Nnenna Nwakanma, Former Policy Director of the Web Foundation. Board Member of I-DAIR.



This inaugural report highlights the urgent need to address the fragile relationship between our democracies, economies and information ecosystems.

The work of the OID is crucial as it provides a comprehensive global assessment of how these ecosystems influence information integrity, political fairness, media freedoms and fundamental rights in general. This report distils extensive research, offering insights into the interplay between Internet access, artificial intelligence, media freedom, and data governance. By examining these interdependencies, the OID's analysis illuminates the significant role that mis- and disinformation play in undermining democratic processes, while also deploying negative consequences on entire sectors of our economies and societies.

This report serves as a landmark contribution by establishing a foundational understanding of the challenges faced by democracies worldwide. It underscores the necessity for policymakers to adopt its key recommendations urgently, particularly in an era where big tech companies dominate information landscapes.

Luca Belli, Professor and Director of the Center for Technology and Society at Fundação Getulio Vargas (FGV) Law School, Rio de Janeiro.



The Observatory is a very important initiative because it brings together perspectives from all continents and many countries and thus helps to develop a more nuanced understanding of policy issues. The Observatory also helps create a transnational network among experts and practitioners who are likely to benefit from each other in the future. Regarding Information Ecosystems and Troubled Democracy, the Observatory is able to demonstrate in a unique way the global reach combined with the regional diversity of challenges we face.

Jeanette Hofmann, Professor at Freie Universität Berlin, Research and co-founding director of the Alexander von Humboldt Institute for Internet and Society (HIIG).

PREFACE



BY ROBIN MANSELL, SCIENTIFIC DIRECTOR
Professor Emerita, Department of Media and Communications,
London School of Economics and Political Science.

What does research in the humanities, social sciences and computer science tell us about the contribution of information ecosystems to the viral spread of mis- and disinformation and about the state of democracy? It has been a great honor for me to lead the International Observatory on Information and Democracy's critical synthesis of research on this question. My collaborators and I have enjoyed a remarkable interdisciplinary journey.

We started this project during a period of renewed technology boosterism – generative AI – and of growing concern about deception, the rise of populism and political polarization, and the gap between the very rich and the disadvantaged, alongside the existential crisis of climate change. This report is about how the transformation of online spaces is connected to everyday life and to the political realm. We were guided in our work by a common position – harms linked to mis- and disinformation need investigation both as symptoms of changes in societies and as amplifiers of these changes. This implies a holistic approach. However, not every piece of research simultaneously examines every aspect of economic, cultural, social and political change in societies or all the impacts of inaccurate information or hate speech. Our response was to recognize the value of a variety of ways of knowing. This meant a readiness to engage respectfully with the multiple theories and empirical research strategies and methods used to investigate our topic.

My reflection on our work is that centering on information (or news media) or technology in research concerned with troubled democracy is unproductive. Such centering is common, but it directs attention away from the power relations that structure discourses and institutionalized public and private action in the world. When these power relations – and how they change – are neglected it is very hard to imagine what a truly democratic information ecosystem could be, or to work out which actors are best positioned to mobilize action to bring it into being.

Readers will find lacunae (especially in capturing insight in the Global Majority World). Inadvertent mis-readings of the evidence we reviewed may be present and feedback is welcome for the future work of the Observatory. My hope is that this report encourages readers to think hard about what the most important questions are for future research. This report reveals why research in an increasingly politicized field of inquiry can yield contradictory results that are strongly contested. It brings to light research on the effects or impacts of mis- and disinformation, but crucially also on asymmetries in power that enable governments and companies to engage in strategies and practices that fail to protect human rights. This report makes clear that a reimagining of ways of structuring and governing information ecosystems is essential if they are to promote the common good.

The analysis in this report helps to penetrate the veil of overhyped technology and the rhetoric that imagines existential threats at every turn. On issues of human safety, security, privacy, cyberwar, or ‘deepfakes’, governing steps being taken now affect everyone. The outcomes are unlikely to be just and equitable without research that produces evidence on why, not just how, harms to adults and children and their communities occur. Such evidence must be sensitive to the diverse contexts and system interdependencies that influence the configuration of information ecosystems and the feasibility (or failure) of inclusive debate within the public sphere. This research synthesis is only a start. It does not provide universally valid grounded truths in answer to our questions, but it maps a terrain of inquiry that is crucial to pursue.

I cannot express enough gratitude to all those who participated in, commented on, and authored this report. We received outstanding support by the Observatory’s Team, especially Iris Boyer whose guidance and encouragement were always timely and much appreciated, and Emma Gruden whose work went far beyond what anyone would expect. On behalf of myself and my collaborators in this project I thank the Steering Committee for its helpful and detailed criticisms of the several drafts of this report.

TABLE OF CONTENTS

Dedication

Foreword iii

Insights from the Steering Committee vii

Preface ix

Table of Contents xi

Executive Summary xv

Acknowledgements xxiii

Chapter 1.

Information Ecosystems and Democracy

1	Introduction	1
2	Setting the Context: Datafication and Democracy	2
3	Positioning the Research Assessment: Concepts and Definitions	6
4	Troubled Democracy and Mis- and Disinformation	9
5	Limitations of the Report	12
6	Chapter Summary and Report Outline	13

Chapter 2.

News Media, Information Integrity and the Public Sphere

1	Introduction	17
2	News Media and Structural Power	18
3	News Media, the Public Sphere and Democracy	22
4	Trust in News Media	24
	4.1. Changing Journalism Standards and News Media Practices	25

4.2.	News Media Trust and Audience Engagement	26
4.3.	News Media Use, News Avoidance and Resilience	29
4.4.	Public Opinion and Polarization	37
5	Strengthening Trust and Resilience to Mis- and Disinformation	41
6	Chapter Summary	42

Chapter 3.

Artificial Intelligence, Information Ecosystems and Democracy

1	Introduction	47
2	AI Systems and Human Rights	49
	2.1 New Technologies – But No New Rights	49
	2.2 Algorithmic Bias and Fairness	51
	2.3 Freedom of Expression and Information	52
	2.4 Privacy Protection	53
	2.5 Democracy and Participatory Rights	55
3	AI Systems and Content Governance	55
	3.1 AI Systems in Content Generation	56
	3.2 AI Systems in Content Moderation and Curation	57
	3.3 AI Systems and News Media	59
	3.4 Use of Generative AI by Mis- and Disinformation Actors	63
	3.5 Countering Mis- and Disinformation	65
4	AI Systems and Democracy	66
	4.1 AI Systems and Mediated Public Sphere(s)	66
	4.2 AI Systems and Societal Resilience and Cohesion	68
	4.3 AI Systems and Social Sustainability	69
	4.4 AI Systems and Environmental Sustainability	70
5	Chapter Summary	71

Chapter 4. **Big Tech Power and Governing Uses of Data**

1	Introduction	75
2	Digital Infrastructure Contestations	75
3	Corporate Data Monopolization and Information Infrastructures	77
3.1	Data Monopolization and Data Dependency	78
3.2	Big Tech Monopolization	81
3.3	Business Models and Mis- and Disinformation	83
4	Towards Democratic Data Governance	86
5	Chapter Summary	88

Chapter 5. **Awareness of Mis- and Disinformation and the Literacy Challenge**

1	Introduction	91
2	Scale and Severity of Mis- and Disinformation	91
3	Public and Policy Maker Awareness of Mis- and Disinformation	93
3.1	Public Awareness of the Impacts of Mis- and Disinformation	94
3.2	Policy Makers' Awareness of Risks and harms	97
4	Literacies for Navigating Information Ecosystems	100
4.1	Ability to Engage Safely Online	100
4.2	Media and Information Literacy	102
4.3	AI Literacy	104
4.4	Effectiveness of MIL and AI Literacy Initiatives	106
5	Chapter Summary	108

Chapter 6. **Governing Information Ecosystems: Legislation and Regulation**

1	Introduction	112
2	Types of Governance Approaches	112
3	Global Governance of Information Ecosystems	114
4	Governance Approaches Applied at Regional and National Levels	116
4.1	Governing Network Infrastructure	116
4.2	Privacy and Data Protection Governance	118
4.3	Governing Digital Platforms	121
4.4	Governing AI Systems	125
4.5	Governing News Media	127
5	Chapter Summary	130

Chapter 7. **Combating Mis- and Disinformation in Practice**

1	Introduction	133
2	Assessing Measures to Combat Mis- and Disinformation	133
2.1	Digital Platform Strategies to Combat Mis- and Disinformation	134
2.2	Digital Platform Strategies to Combat Mis- and Disinformation	137
2.3	Effectiveness of Mis- and Disinformation Countermeasures	138
3	The Challenges of Information Ecosystems Governance	142
3.1	Governing AI Systems: Accountability and Transparency	142
3.2	Governing News Media and Press Freedom	144
3.3	Governing Data in Corporatized Information Ecosystems	145
4	Human Rights and Mis- and Disinformation Countermeasures	147
5	Public Appetite for Combating Mis- and Disinformation	149
6	Chapter Summary	152

Chapter 8.

Towards Data Justice in Information Ecosystems

1	Introduction	155
2	Strengthening Deliberation and Democracy	156
2.1	Corporate Power and Interests	156
2.2	Data, AI Systems and Discriminatory Bias	159
3	Alternative Data Governance Practices	161
3.1	Individual Digital Self-Defense Strategies	164
3.2	Public Interest Alternative News Media	165
3.3	Community Collaborative Strategies	166
4	Chapter Summary	172

Chapter 9.

Conclusion: Information Ecosystems and Troubled Democracy

1	Introduction	175
2	Principal Thematic Insights	176
2.1	Human Rights	177
2.2	Contesting Data Monetization	177
2.3	Exclusion and Inequitable Inclusion	178
2.4	Transparency and Accountability	180
3	State-of-the-Art Research and Future Directions	181
3.1	A Eurocentric/Western Research Bias	182
3.2	Conceptual Framings	182
3.3	Research Design and Methods	183
3.4	Researcher Access to Data	184
3.5	Research Independence	185
4	Chapter Focus and Organization	186
5	Limitations of the Report	190
6	A Final Word on what should be done	191
6.1	A Role of civil society organizations	191
6.2	Guidance for policy makers	192
6.3	Guidance for big tech companies	194

Appendix: Methodology	196
------------------------------------	-----

Bibliography	211
---------------------------	-----

List of Tables

• Table 1.1: Distinguishing between mis- and disinformation	10
• Table 2.1: Tensions between digital platforms and news media organizations	20
• Table 2.2: Platform roles in the weaponization of information – selected country examples	36
• Table 3.1: AI system use cases in the news media gatekeeping process	61
• Table 6.1: Governing information ecosystems	114
• Table 7.1: Selected mis- and disinformation countermeasures	134
• Table 8.1: Corporate datafication resistance strategies and tactics	163
• Table A1: Definitions of source quality	203

List of figures

• Figure 1.1: Information ecosystems and the public sphere	7
• Figure 1.2: Polluting the information ecosystem	10
• Figure 2.1: Building trust in news	42
• Figure 3.1: Illustration of user engagement	53
• Figure 3.2: Example of realistic AI-generated face using the 2020 algorithm StyleGAN2	63
• Figure 3.3: Deepfake image of Donald Trump generated using Stable Diffusions	65
• Figure 4.1: Generative AI promotion	80
• Figure 4.2: Large Language models for efficiency	80
• Figure 4.3: Map of mis- and disinformation campaigns in West Africa, and state sponsorship	85
• Figure 5.1: Understanding of 'AI'	94
• Figure 5.2: Proportions of respondents indicating they have ever used ChatGPT, by age	95
• Figure 5.3: Country averages of proportion of 18–24-year-olds using social media weekly, 2014–2023	96
• Figure 5.4: Spread of posts on X on Southport UK murder, 2024	98
• Figure 5.5: Six cues for snap judgements about what to trust	101
• Figure 5.6: Non-exhaustive MIL competencies	104

• Figure 6.1: Simplified view of the governance of information ecosystems	112
• Figure 7.1: Network of fact-checking organizations, 2014–21	135
• Figure 7.2: World Press Freedom Index, 2024	144
• Figure 8.1: Information ecosystems – alternative data governance approaches and resistance strategies	162
• Figure 9.1: Principal thematic insights	176
• Figure 9.2: State-of-the-art research assessment	181
• Figure 9.3: State of academic freedom, 2023	185
• Figure A1: Number of citations in the report by publication year (N=1,664)	203
• Figure A2: Number of citations in the report by publication year (2000–2014, N=110)	204
• Figure A3: Number of citations in the report by publication year (2015–2024, N=1,508)	204
• Figure A4: Most common publication types (N=1,664)	204
• Figure A5: Number of publication types by year, 2000–2024 (N=1,618)	204
• Figure A6: Types by year, bar chart, 2000–2024 (N=1,618)	205
• Figure A7: Types by year, cumulative, 2000–2024 (N=1,618)	205
• Figure A8: Number of types of publications, by regional focus (N=1,664)	206
• Figure A9: Share of types of publications, by regional focus (N=1,664)	206
• Figure A10: Share of types of cited publications by regional focus (N=1,664)	206
• Figure A11: Number of books and book chapters by most frequently occurring publishing company	207
• Figure A12: Most frequently cited journals in the report	207

How to cite this report: Mansell, R., Durach, F., Kettemann, M., Lenoir, T., Procter, R., Tripathi, G., and Tucker, E. (2025) Information Ecosystems and Troubled Democracy: A Global Synthesis of the State of Knowledge on New Media, AI and Data Governance. International Observatory on Information and Democracy. Paris.

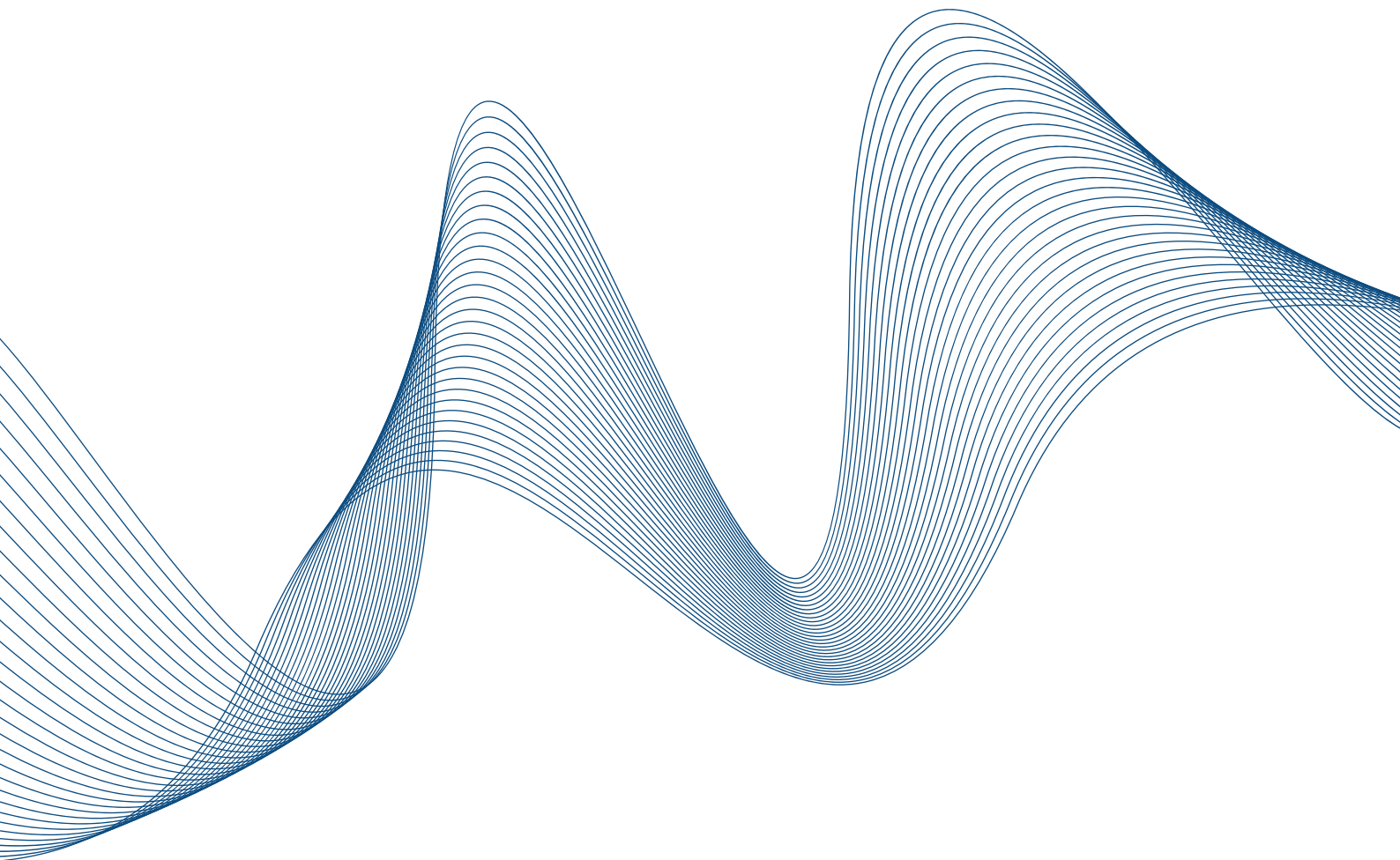




EXECUTIVE SUMMARY

INFORMATION ECOSYSTEMS AND TROUBLED DEMOCRACY

A Global Synthesis of the State
of Knowledge on News Media,
AI and Data Governance



1 Introduction

Democracy is troubled. There is no dispute about this. What is controversial is the role of information ecosystems in contributing to the fragility of democracy and to the viral spread of mis- and disinformation. The V-Dem Institute reports that the level of democracy enjoyed by the average person globally in 2023 had declined to 1985 levels – 5.7 billion people living in autocracies.¹ Distrust in online information is widely seen as an ‘information crisis’. Research on digital news, covering six continents, found that people’s concerns about what is real and what is ‘fake’ news had risen on average to 59% of those sampled, and to as high as 72% in the United States and 81% in South Africa.² When the World Economic Forum interviewed experts in 2024, they placed AI-generated mis- and disinformation produced by domestic and foreign actors at the very top of a list of global risks.

Big tech company business models entice children and adults online to allow the extraction of data, which they then monetize for profit. This is facilitating the viral spread of mis- and disinformation and hate speech. While mis- and disinformation have always been an issue, information manipulation and distribution are now supported by artificial intelligence (AI) tools and algorithms. At the same time, there is uncertainty about how to ensure that international human rights commitments are met, and that information ecosystems foster democratic debate in the public sphere. The tensions between efforts to address mis- and disinformation and measures to address human rights commitments are troubling to democracy, and solutions seem elusive.

This report is a critical review of state-of-the-art research in three areas, with a cross-cutting theme of mis- and disinformation: **media, politics and trust; artificial intelligence, information ecosystems and democracy; and data governance and democracy**. The role of information ecosystems in both the Global North and the Global Majority World is assessed, with a focus on their relationship with information integrity (the quality of public

discourse), the fairness of political processes, the protection of media freedoms and the resilience of public institutions.

The end of the International Observatory on Information and Democracy’s first year-long critical review of state-of-the-art research coincided with the publication of the United Nations’ *Global Digital Compact* in September 2024. This commits member states to ‘promote diverse and resilient information ecosystems’. Our analysis is based mainly on academic publications supplemented by reports and other material from different disciplines and regions (1,664 citations selected from our bibliographic database, with more than 3,095 entries screened before inclusion). This report focuses on questions set by the Observatory’s Steering Committee (see Appendix: Methodology). It is not intended to prescribe specific actions for policy makers; rather, it showcases what we can learn from landmark research on the often intractable challenges posed by rapid changes in information and communication spaces. These spaces can be viewed as ‘information ecosystems’.

We understand ‘information ecosystems’ as systems of people, practices, values and technologies configured in social, cultural, political and economic contexts. The interdependencies of these components are complex and they involve structural and power relations among multiple actors. How these operate in a specific context is what conditions the integrity of information and the possibilities for informed participation in the public sphere. Information integrity is understood as ‘access to relevant, reliable and accurate information and knowledge’ following the United Nations *Global Digital Compact* definition. Democratic participation requires information integrity and the existence of inclusive, open, safe and secure digital spaces, where there is tolerance and respect.

Shortcomings in either information integrity or the safety, security and inclusiveness of digital spaces are understood to undermine the vitality

¹ V-Dem Institute. (2024). *Democracy Report 2024: Democracy Winning and Losing at the Ballot*. Varieties of Democracy Institute.

² Newman, N., et al., (2024). *Digital News Report 2024*. Reuters Institute for the Study of Journalism, University of Oxford.

of the public sphere and, in some cases, even its existence. In this report, research on the reality of asymmetrical power relations between big tech companies, states and publics receives particular attention with a view to revealing the agency of both individual actors and institutions to address these asymmetries. It is acknowledged that there are definitions of information ecosystems that place greater emphasis theoretically on the indeterminacy of the evolutionary dynamics of ecosystems and therefore on the uncertainty of outcomes of interventions aimed at reducing power asymmetries.

In this report, we do not assume that the design, deployment, beneficial and harmful uses of digital technologies are dictated by technological change; rather, information ecosystems are understood to be a result of human decisions and actions. Our interest is in what research reveals about the interdependence of changing information ecosystems with the public sphere and democracy. Harms associated with mis- and disinformation are treated as both symptoms of complex changes in society and as important amplifiers of these changes.

2 Structure of the Report

Our critical review of research begins in Chapter 1 with a discussion on the core themes and definitions of the key concepts. The rest of the report critically introduces readers to existing research with a focus, first, on each of our three core themes on media, AI systems and data governance (Chapters 2–4). We then turn to research that cuts across these themes to focus on public understanding of mis- and disinformation and literacy training, governance of information ecosystems, practices aimed at combating mis- and disinformation, and strategies for achieving data justice (Chapters 5–8). Finally, Chapter 9 concludes

with a synthesis of the key research insights, future research directions and guidance for policy makers and big tech companies.

The main issues and principal questions addressed in each chapter are now summarized.

News Media, Information Integrity and the Public Sphere (Chapter 2). Here we look at what research tells us about changes in legacy and online news media, and what can be done to promote information integrity and a democratic public sphere. What are the salient changes in news media industry market structures and power relations between news media organizations and digital platforms? What is the relationship between news media, a healthy public sphere and democracy? How is trust in news media associated with political polarization? What strategies are available to the journalism profession and other actors to build trust in the news?

Artificial Intelligence, Information Ecosystems and Democracy (Chapter 3). This chapter focuses on the properties of AI systems (machine learning algorithms) and the consequences of their being embedded in content governance systems. How does ‘AI’ operate in ways that affect information integrity? What is the relationship between AI systems and internationally protected human rights? What are the interdependencies between AI systems, the use of automated tools and democratic processes?

Big Tech Power and Governing Uses of Data (Chapter 4). Here attention turns to the power of big tech companies, and approaches to governing data extraction and use (processes of datafication). What is the appropriate

role of data and digital infrastructures within political communities? How are data aggregation and AI systems changing the way people build, share and receive information? How do these big tech strategies and practices influence political deliberation?

defining and implementing these approaches? In what ways is human rights protection jeopardized by efforts to curtail mis- and disinformation? What do we know about the public's appetite for interventions to moderate online mis- and disinformation?

Awareness of Mis- and Disinformation and the Literacy Challenge (Chapter 5).

This chapter examines research on people's knowledge about mis- and disinformation, as well as literacy training initiatives aimed at enabling people to protect themselves from online harms and to distinguish inaccurate from accurate information. How aware are the public and policy makers of the risks and harms of mis- and disinformation? What are the approaches to media and information literacy (MIL) and AI literacy, and are they effective?

Towards Data Justice in Information Ecosystems (Chapter 8). Research

indicates that the monopolistic power of big tech companies in data extraction and monetization leads to harmful discrimination and exclusions. Why do corporate strategies and practices lead to epistemic injustice? What strategies and tactics are individuals and communities developing to resist the extractive features of the data economy?

Governing Information Ecosystems:

Legislation and Regulation (Chapter 6). Here we examine selected legislative and regulatory tools that aim to mitigate the harms of mis- and disinformation and to govern how big tech companies operate. What governance approaches are available? What approaches are being promoted at the global level? What insight can we draw from the variety of legislative, regulatory and judicial approaches at national and regional levels?

Conclusion: Information Ecosystems and Troubled Democracy (Chapter 9). In this

chapter themes that emerged from our analysis across the report (issues relating to human rights, contesting data monetization, exclusion and inequitable inclusion and transparency and accountability) are discussed. Prominent characteristics of the research we reviewed are presented (a Eurocentric/Western bias, inconsistent conceptual framings and a wide variety of research designs and methods, limited access to research data and the challenges of securing research independence). This concluding chapter highlights future directions for research, provides a summary of findings by chapter, and distills guidance for policy makers and big tech companies (no specific recommendations are made as this was not the purpose of our review).

Combating Mis- and Disinformation in Practice (Chapter 7). This chapter turns

to specific measures to combat mis- and disinformation by civil society organizations and governments. What content governance approaches are used to combat mis- and disinformation? What are the challenges in

3 What Can We Learn from a Critical Review of State-of-the-Art Research?

We found broad agreement that states have a duty to protect **human rights and fundamental freedoms**. A consistent emphasis is the need to differentiate between normative goals and principles articulated at a global level, and how these are translated into practice in different contexts. There is a tension between research favoring incremental risk mitigation strategies and research concluding that human rights protections are incompatible with big tech data extraction practices.

The need for research on how international human rights law is interpreted and applied at regional and country levels was emphasized repeatedly.

Data monetization for profit is a prominent topic. Big tech business models are shown to drive developments on the infrastructure layer of information ecosystems – for example, network neutrality policies and ‘zero-rating’ contracts – and on the service applications layer – for example, destabilizing news organizations’ finances and facilitating the weaponization of information. A common theme is that policies favoring the data dependency of private and public organizations, as well as individuals, pre-empt meaningful political deliberation on issues such as rights to data ownership, what role data should have in the private and public sectors, and what contexts require the minimization or prohibition of data production.

Strategies that aim to counter harmful exercises of power would benefit from research aimed at exposing how big tech business models make them attractive targets for mis- and disinformation

campaigns. Research is also needed on the new competencies and enforcement mechanisms required for combating harms to a diverse public sphere.

Research on **exclusions from and inequitable inclusions in information ecosystems** at local, national and regional levels is not as prominent in the literature we reviewed as the first two themes. Many studies conducted in the Global North do not acknowledge that (meaningful) internet connectivity is absent for many in the Global Majority World.

There were 5.4 billion internet users in 2024, and 2.6 billion people with no access; in low-income countries 20.9% of people use the internet; in high-income countries the figure is 90.5%.³

Such studies, for example, do not sufficiently recognize that globally news media systems are subject to a variety of ownership and regulatory regimes; that content governance measures often suppress debate that is critical of authorities; or that AI systems impact communities of color, women, religious minorities and LGBTQ+ people in harmful ways. When these inadequacies are highlighted, it is found that big tech companies are involved in replicating and exacerbating inequalities and injustices. The *Global Digital Compact’s* ambition is to tackle exclusions and inequitable inclusions. In the academic literature reviewed in this report, evidence of practical steps to ensure the Global Majority World is not treated as a passive recipient of Eurocentric/Western ideas is lacking.

People in the Global Majority World need to be heard, and barriers to their participation in decisions need to be reduced, so that elite Global North knowledge is not the unquestioned guide to governing information ecosystems and the public sphere.

³ ITU. (2024). Statistics ITU; ITU (2024). *The ICT Development Index 2024: Measuring Digital Development*.

Transparency and accountability issues are discussed, but there is a tension between research claiming that governance of information ecosystems is too permissive or that it is not permissive enough. Where governance is found to be too permissive, this is because economic self-interest is given priority without sufficient attention to rights protection. When governance is found to be not permissive enough, it is most often because states are unduly suppressing speech. The governance of those harvesting and selling data needs reinforcing. However, in the Global Majority World, there are concerns about the feasibility of holding distant actors to account, with little clarity about the interventions that would be most effective in mitigating the harms of mis- and disinformation. There is a strong emphasis on the need to promote the transparency of AI systems and independent audits.

To hold big tech companies and governments accountable, accurate information needs to reach a wide range of stakeholders. Actors who question mis- and disinformation governance practices should be neither criminalized nor marginalized.

Research has addressed **media and information literacy (MIL) and AI literacy training** as a means to help children and adults keep themselves safe from harmful information. This work focuses on curricula, training and funding, but literacy issues also appear in connection with debates about transparency and accountability. A public that is better informed about factors that facilitate illegal and harmful information is more likely to demand that big tech companies and states are held to account, to insist on transparency (as far as possible) of algorithmic systems, and to argue for human oversight of algorithmic decisions.

Literacy initiatives should not, however, be a stand-alone answer to mis- and disinformation problems. There is little systematic evidence of experience of literacy initiatives globally, and over time, and there is less research on children's literacy than on those of adults.

Our analysis revealed several key characteristics of state-of-the-art research on information ecosystems and the challenges of mis- and disinformation.

There is a clear **Eurocentric/Western bias** towards research in and on the Global North, with the problems of mis- and disinformation and approaches to mitigating harms studied disproportionately in the United States and other Western countries. Research on companies – small and large – that produce discriminatory outcomes as the result of datafication focuses on relatively few large companies. There are few in-depth assessments of experience around the world, apart from some comparative survey studies.

This research bias must be addressed if the views of individuals and organizations in the Global Majority World working on mis- and disinformation are to inform policy, in both the Global Majority World and at the international level.

The **conceptual framing** of issues in research cited in our report relies on multiple definitions. Even if there is some consistency in defining concepts in policy documents, meanings differ across disciplines and in different regions/countries. There are tensions between whether the object of interest is an information ecosystem or the public sphere. 'Information integrity' is criticized as being too open to interpretations of what is good or 'polluting' information, and for neglecting the history of research on propaganda and the public sphere.

Building bridges between the humanities, social sciences and sciences could help to resolve inconsistencies, but it is important to recognize that variety is inevitable given diverse information ecosystems. Conceptual framings would benefit from joined-up research on the public sphere and democracy, including work on securitization and the socio-economics of online labor markets.

There is also a tension in **research design and methods**. For example, some research aims to detect causal links between mis- and disinformation and algorithmic personalization systems. Other research aims to reveal power asymmetries that underpin commercial datafication systems. Both offer an insight into mis- and disinformation, echo chambers and political polarization. However, fewer studies examine the interdependent relationships between components of information ecosystems that sustain asymmetrical power relationships, including the monopolistic power of big tech companies.

Research on mis- and disinformation relies extensively on quantitative experimental or quasi-experimental designs or is based on surveys. Qualitative methods can help to reveal how power disparities influence choices about the design and deployment of digital technologies and the agency of individuals and groups – why, for example, people value online echo chambers, how AI systems are understood to operate, or why trust in news media and perceptions of the trustworthiness of news media organizations varies as much as it does across countries.

Multidimensional (holistic) research is needed on factors that enable the creation and circulation of mis- and disinformation. This research should pursue research designs and methods that provide an insight into the affordances of technology and the practices of states, companies and other actors.

Researcher access to data is discussed as a problem throughout the literature, indicating an urgent need for safe harbors for researchers, as well as clear data disclosure policies. The importance of securing the independence of researchers and their institutions in the face of efforts to suppress research that is deemed politically sensitive, or that questions the claims of companies, is emphasized. While clearly needed in relation to research on the role of mis- and disinformation in elections, independent research is also needed in areas such as the responsible development of AI systems.

Monitoring the independence of researchers and their institutions, as well as the impacts of corporate and government funding, is essential.

4 Detailed Insights and Future Directions for Research

This section highlights insights from our analysis of research on each of the core themes, and points to additional future directions for research.

4.1. NEWS MEDIA, POLITICS AND TRUST

We focused on changes in the legacy and online news media industry, and how these are associated with the structure of markets, political processes and trust in news media and political institutions (Chapter 2).

The rise of monopolistic digital platforms

owned by big tech companies was shown to be threatening the viability of news production as well as influencing news consumption. The extent of news media dependence on the platforms and pressures on legacy news finances was shown to vary by country, type of organization, print/broadcast versus online organization, and by news organizations' strategies to address their audiences. The news media industry is in crisis in many countries, but the reasons and outcomes differ. In some countries news media concentration is reducing the diversity and plurality of news content; in others, financial instability is leading to news deserts. A lack of sustainable financing is putting pressure on journalists and their editorial processes, and influencing public perceptions of the trustworthiness of news organizations.

The platformization of news has led to a power asymmetry between news media organizations and digital platforms that must be addressed.

Measures are needed to tackle the monopolistic power of big tech companies when it leads to harm to individuals and unhealthy information ecosystems. Measures are needed specifically to require disclosure of revenue and online traffic, so that the value of news hosted by platforms can be established. Smaller news organizations need support to achieve greater bargaining power in their dealings with platforms.

We reviewed research on who consumes the news, whether they trust it, and how news exposure influences attitudes and behaviors. News media trust was shown to depend on variables such as age, gender, education, ideology and partisanship and socio-economic status, with trust varying significantly across countries.

A total of 40% of respondents self-reported trust in news most of the time: Finland had the highest overall trust, at 69%; United States, 32%; France, 31%; Argentina, 30%; Greece, 23%; Hungary, 23%; there was little evidence that upcoming elections at the time of the survey impacted on indicators of trust.⁴ However, 87% of survey respondents in 16 countries reported being worried about the impact of disinformation in upcoming elections in 2023.⁵

In both democratic and autocratic countries, interest in and knowledge about politics was shown to influence trust in news and in the trustworthiness of news media organizations. The relationship between political interest and news media trust was shown to be becoming stronger over time (in some countries). Variations in self-reports of survey respondents across countries are striking and, where trust is

declining, this seems set to continue. Trust in news media and in political institutions is declining in some countries in the Global North, and in some countries in the Global Majority World it remains high. As social media use increases, news exposure also grows, and evidence shows that people access news even if they do not trust it. Evidence indicates declining overall regular engagement with news and that people often choose not to engage at all. This complicates interpretations of the implications for the public sphere.

Research on the effects of mis- and disinformation on media trust focuses principally on individual effects. Research is needed on the agency of online participants/audiences and their capacity to engage in critical thinking about information and news specifically.

A total of 22% of people across 46 countries in 2023 were active participants with online news; 47% were not participating at all. 39% report avoiding news, up 3% on the previous year's average, with the biggest increases in Brazil, Finland, Germany and Spain.⁶

News avoidance is shown to be increasing, although this is uneven across countries. Together with resilience to mis- and disinformation, use or avoidance of news is shown to be influenced by similar factors that influence news media trust. Studies aiming to identify the effects of news media exposure on attitudes, and behavior generally, acknowledge that the effects they detect are likely to vary with context. Research highlights that mistrust in information may lead to more informed public debate, but that where mistrust leads to news avoidance, this isolates people from public life. Research confirms a perceived rise in exposure to 'fake' online news, and that cognitive biases can lead to overconfidence in abilities to detect mis- or disinformation.

⁴ Newman, N., et al., (2024). *Digital News Report 2024*. Reuters Institute for the Study of Journalism, University of Oxford.

⁵ Ipsos & UNESCO. (2023). *Survey on the impact of online disinformation and hate speech*. Ipsos and UNESCO.

⁶ Newman, N., et al., (2023 and 2024). *Digital News Reports 2023 and 2024*. Reuters Institute for the Study of Journalism, University of Oxford.

Research on whether viral mis- and disinformation are principal causes of political polarization is challenging to interpret. Online echo chambers do not appear to be solely attributable to online personalization systems, and some studies emphasize that a minority of people consume mis- and disinformation. Some research finds no evidence of direct effects of mis- and disinformation on political polarization or voting behavior. Others show increases in the likelihood of believing stories that favor preferred candidates, amplification of negative emotions or perverse effects of efforts to raise awareness about mis- and disinformation if this leads to distrust in legitimate information.

There is a strong bias towards evidence in the Global North. Studies often rely on short time periods, typically use experimental methodologies and examine a limited number of digital platforms (partly due to difficulties in accessing platform data).

Evidence on the effects of mis- and disinformation on attitude polarization and voting behavior leaves unanswered questions about the effects of different types of information, for example conspiracy theories or lies propagated by politicians. Although exposure to like-minded political content can be associated with political polarization, partisan online echo chambers were found in some research to be smaller than typically assumed in policy debates. Evidence also shows that filter bubbles and echo chambers can have positive effects if they provide a safe space for marginalized groups. The weaponization of information is facilitated by AI systems, including algorithmic personalization systems, and the potential for misuse is likely to increase. Personalization tools enable election campaigns to target voters and disseminate false information with the aim of manipulating attitudes and voting behavior. These systems are used to produce and curate content to favor emotionalizing content, increase online engagement or reward certain social and political groups.

Although information is wielded as a weapon by foreign and domestic actors, there is a bias towards researching far-right groups that do the

bidding of foreign powers, rather than domestic actors. Comparative work is also scarce, despite the Global Majority World experience displaying evidence of weaponization of information related to internal politics. Overall, it seems that automated personalization does not provide a complete explanation for the emergence of echo chambers, and some research emphasizes that who generates mis- and disinformation and why is as important as its effects on political outcomes.

Research is needed on actors (state/private and foreign/domestic) who generate and share mis- and disinformation and their motivations. Also needed are deeper investigations of the contexts in which news is produced (liberal democratic/autocratic), media ownership and market structures, presence or absence (and investment in) of public service media and harms experienced by journalists who try to report accurate news.

4.2. ARTIFICIAL INTELLIGENCE, INFORMATION ECOSYSTEMS AND DEMOCRACY

AI systems, including large language models (LLMs), raise issues for human rights protection and for content governance and democracy (see Chapter 3). Definitions of 'AI' are now being agreed internationally in policy contexts, but in popular discourse, AI is often referred to misleadingly. There is also a variety of definitions in the research literature. There is not an AI, but different machine learning (ML) technologies involved in processes related to information creation, retrieval, synthesis, presentation and governance. It is important to be specific about what AI tools are being discussed in research. In this report we refer to AI systems or to specific types, such as LLMs or generative AI (GenAI).

Internationally protected human rights and fundamental freedoms, including media freedom and freedom of expression, are fully applicable to the production and use of AI systems, but it is important to note that not all countries are

fulfilling their obligations. Some studies insist that AI systems are neutral or can be made neutral, but the weight of evidence is that biased outputs of AI systems are the inevitable consequence of biases in the data on which they are trained. It is clear that algorithmic fairness requires comprehensive strategies to improve data diversity, enforce transparency and ensure that regular algorithmic audits are conducted. It also requires that commitments to responsible and ethical AI use are fulfilled.

Research is needed to identify known biases and to scan for potential biases, and to mitigate them as far as possible by changing AI systems operation and by ensuring that a human is involved in the uses to which outputs are put. Research must focus on whether human rights commitments are being upheld in the wake of global promotion of trustworthy AI for sustainable development.

If AI systems are to meet expectations for fairness and to be consistent with internationally recognized human rights law, research must focus on AI systems explainability and best practices for achieving accountability of automated content governance. Research reveals a lack of accountability of these systems, and weak evidence on the transparency of the training and deployment of automated content governance tools. These systems need to be evaluated using real-world data as well as field research (not only experiments) using quantitative and qualitative methods. It is also clear that no single content moderation technique will be acceptable to every online participant. In addition, safeguards are needed to prevent the platforms using these systems to intensify societal inequalities, contributing to the declining quality of information.

Discussion about the contribution of AI systems to the benefit or detriment of information ecosystems and the public sphere must be as inclusive as possible. The growing 'AI divide' requires thorough investigation, along with the obstacles

that prevent people in the Global Majority World from participating in decisions related to developing and implementing AI systems.

There is an urgent need to deepen understanding of how the organizational principles – norms and rules – of private information and communication systems, along with their algorithmic designs, affect society. This knowledge must be leveraged to hold those who deploy AI systems accountable for their decisions. AI systems are only one factor in societal transformation, but decisions taken in their design and operation can diminish or enhance societal resilience and cohesion. LLMs also demand vast amounts of data and energy-intensive training processes.

Research encompassing the whole life cycle of AI systems development, including environmentally responsible innovation, in diverse use and country contexts is urgently needed. This requires access to corporate data and the mechanisms of interlinked personalization systems that are rarely shared with researchers.

4.3. BIG TECH POWER AND GOVERNING USES OF DATA

Research on data governance and datafication (turning offline action into online quantified data for tracking and predictive analysis) provides an insight into the strategies and practices of big tech companies and the structure of power relations in commercial markets (Chapter 4).

Two main types of monopolistic activity are of concern: the monopolization of user data (i.e., all the data produced about us), which makes money for companies by converting information seekers into 'information products' offered for sale to advertisers; and the monopolization of knowledge (i.e., data organized as usable insights), which occurs when data resources (including public data resources) are converted into private assets. Big tech companies do not limit their data collection to

the data they extract, but also develop symbiotic or parasitic relations with other less well-known companies that amass, analyze and sell data. This leads to the dependency of individuals, economic sectors and multiple spheres of public and private life that are mediated by the choices of these companies.

Data-intensive algorithmic products, marketed as 'AI', are shown to pose significant threats to information ecosystems and to democracy because data and information are structured in ways that few understand or have control over. This affects their ability to resist manipulation and to deliberate with others about the common good.

These forms of datafication give rise to numerous forms of digital dependency. Data infrastructures are shown to be pervasive and largely invisible, yet also determinative.

Many factors are helping to create conditions in which the data-related features of digital infrastructures are proliferating. Digital infrastructures are being imposed because of under- or unregulated corporate activity alongside opaque government procurement processes. In the face of a GenAI 'arms race', discussion around data governance is at an all-time high. Research demonstrates how legislation and frameworks that govern uses of data foster the amplification of mis- and disinformation, and that companies are creating *de facto* data governance frameworks for data use that ignore the amplification of mis- and disinformation. The lack of robust, and robustly enforced, rules about which public and private actors can do what with respect to data is a primary reason for these negative consequences.

Data governance is being addressed in relation to the privacy, security and integrity of data, but there is strong political pressure from within civil society to think about the role of data governance as a lever for restructuring markets to protect people against human rights infringements and concentrations of power and wealth that are inconsistent with democracy. The roles of data, data-dependent

digital infrastructures, data markets and companies in the data business are being questioned. This must be part of any democratic digital policy-making project.

This questioning must seek to preserve and promote the capacity of diverse communities to take up such questioning outside formal policy-making spaces. This questioning is necessary not only for democracy, but as democracy.

Common approaches to data governance (including AI governance) focus on protecting security (individual and/or state), property and dignity/autonomy, and more robust enforcement might improve outcomes for individuals and communities. However, these frameworks are shown in the critical literature to be failing to provide a basis for contesting datafication itself. Existing data governance frameworks devised by national regulators, multilateral bodies, companies and multistakeholder organizations are not sufficient to protect most kinds of data from being acquired by large companies to generate revenues and amass political and economic power. Research also shows that individual and collective dependencies and inequities resulting from datafication are experienced differently around the world.

Research must go beyond analysis of the impacts of datafication and AI systems on individuals. A broader range of impacts of datafication in people's lives needs to be documented if meaningful political deliberation about fundamental human rights is to be possible. Developing new data governance frameworks must be a collective effort, involving governments, big tech companies, civil society and political actors.

4.4. MIS- AND DISINFORMATION AWARENESS AND LITERACY CHALLENGES

Chapter 5 examined evidence on the actual scale of mis- and disinformation and public awareness of risks and the severity of its harms, as well as media and information literacy (MIL) and AI literacy initiatives aimed at enabling adults and children to keep themselves safe online.

It is impossible to provide a single or even very meaningful quantitative measure of the scale of mis- and disinformation because of difficulties in collecting and analyzing data that reflects people's online experiences. Evidence on the scale and severity of harms associated with mis- and disinformation comes mostly from surveys and experimental research. Large-scale studies are limited to a few platforms and largely centered on the United States. Privacy protection, ethical issues and big tech company restrictions on access to data create measurement challenges. The data access situation is changing, but the problems are greater for researchers in the Global Majority World than for those in Global North where there are moves to secure better data access for research.

Investment in public data infrastructures for research is essential to enable research that can guide policy and offer insights into the best measures to combat mis- and disinformation.

'The sheer vastness and diversity of online experiences makes meaningful measurement a challenge requiring investment and innovation. The scale and variety of online platforms, and algorithmic personalisation of content, means that there is essentially an infinite number of possible user journeys, making it hard to arrive at both meaningful summary insights as well as fine-grained assessments of particular issues'.⁷

Our analysis confirms a substantial variability in the public's and policy makers' understanding of the threats and impacts of mis- and disinformation, and the role of algorithms and digital platforms in rights protection and democracy. The public's awareness of mis- and disinformation in their daily lives varies, and research confirms that awareness is not the same as an ability to spot inaccurate information. People who are active online may or may not believe they have agency or control over what they do online, and their beliefs may or may not be justified in practice.

Research also indicates that public awareness of the presence of AI systems in their lives and whether they should trust it to make decisions varies across countries. There is too little systematic research on public awareness of differences in people's abilities to avoid mis- and disinformation. Research focusing on differences in this area does not focus enough on connectivity issues, finance, social networks in the offline world or the wider political environment, although there is research on individual characteristics.

A total of 66% of people surveyed thought AI would dramatically affect their lives in the next 3–5 years; 67% reported a good understanding of what AI is.⁸

The promotion of **media and information literacy (MIL) and AI literacy** aims to enable people to protect themselves from online harms linked to online mis- and disinformation. Evidence indicates that literacy training is commanding greater attention now that it is being coupled with AI literacy. MIL and AI literacy appear to be effective means of tackling risks and harms when training is audience- or user-centered. The evidence shows that adults and children with critical literacy skills are likely to be better able to differentiate between legitimate and other sources of information, although sophisticated tools for creating 'deep fakes' are making this harder.

⁷ Faculty. (2021, p. 2). *Automated approaches to measuring online experiences: Executive Summary*. Faculty plc Report for Ofcom.

⁸ HAI. (2024). *AI Index Report 2024 – Artificial Intelligence Index*. Stanford University Human Centered Artificial Intelligence, based on a 2023 Ipsos survey.

It is also clear that AI literacy training for policy makers and digital service designers must include knowledge about the fundamental AI principles of how these systems operate at all stages of AI systems development and deployment. Although some evaluations of literacy training show improvements in how people engage with online information, the duration of effects is unclear, and funding and capabilities for training are not consistently available across countries. Standardized MIL and AI literacy conceptual frameworks and methodologies are needed to advance the evaluation of literacy initiatives. These may benefit from government, private sector or civil society partnerships to promote training.

Research is needed on the efficacy of literacy initiatives for children and adults to assess whether they are better able to discern the accuracy of information over time and to keep themselves safe from harm after training. It is also clear that these initiatives must be complemented by measures that address societal interests in healthy information ecosystems, the value of the public sphere and the usefulness of governance measures.

4.5. GOVERNING INFORMATION ECOSYSTEMS

Research on legislative and regulatory measures (Chapter 6) and on strategies and practices intended to combat mis- and disinformation (Chapter 7) was examined.

Between 2016 and 2022, 91 laws were enacted or amended to deal with misleading information; from 2011 to 2022, a total of 105 new laws or reinforcement of older laws were put in place to combat mis- or disinformation. In the case of AI policy specifically, since 2016 an estimated 800 AI policy initiatives have aimed to tailor AI governance to country

conditions in a way that respects human rights and results in transparency and accountability.⁹

Approaches to governing information ecosystems include corporate self-regulation, state-industry co-regulation and direct state intervention. Research shows that countries are at different stages of implementing legislation and enforcing regulations, and that evidence of their effectiveness is uneven. This applies to rules and norms for corporate data extraction practices, data storage and privacy protection, as well as regulation of digital platforms, AI systems and news media.

Legislation and regulation clearly do not translate automatically into effective enforcement of measures for preventing or mitigating mis- and disinformation harms.

Measures specifically aimed at countering mis- and disinformation are shown to rely on AI-based tools and methodologies, but these are not yet adequate for meeting the challenges of the scale and variety in online platform and user experience. It is apparent that the capabilities of AI systems to tackle mis- and disinformation lag behind AI system capabilities to create these kinds of content. Technical solutions to detect mis- and disinformation are not widely tested beyond laboratory experiments.

The need for investment in real-world testing of the effectiveness of measures to counter mis- and disinformation is urgent.

Comparative studies indicate that the effects of mis- and disinformation countermeasures depend on the type of intervention and information. Research also documents that some measures are used to silence legitimate criticism of the state. Democracies with higher levels of press freedom tend to take a

⁹ Lim, G., & Bradshaw, S. (2023). *Chilling Legislation: Tracking the Impact of "Fake News" Laws on Press Freedom Internationally*. Center for International Media Assistance and Roberts, T., & Bosch, T. (Eds.). (2023). *Digital Citizenship in Africa: Technologies of Agency and Repression*. Zed Books.

holistic approach that focuses on the integrity of the election process, news media diversity and education. Authoritarian regimes, by contrast, are shown to prefer vague responses, allowing them to repress criticism. AI systems and automated tools for combating mis- and disinformation in many instances, either lack regulation or are being used in ways that violate human rights, since big tech companies have the power to decide whether to suppress or amplify information.

Combating mis- and disinformation, including fact-checking methods, needs to be anchored in human rights principles and the rule of law. Requiring ‘meaningful transparency’ and ‘interoperable transparency’ are potential ways to achieve this.

A robust public sphere depends on media freedom, but few countries are achieving ‘good’ press freedom status.

Only 4.4% of countries (eight countries) in the World Press Freedom Index 2023 were ranked as a ‘good’ environment for journalism, down from 14.4% in 2013. Since 1993, 1,701 journalists have been killed, with 50% of these deaths occurring outside conflict zones.¹⁰

Evidence shows that combating mis- and disinformation by regulating the news media can backfire if used as a pretext to consolidate state power and control over information flows. Treating news media as a ‘public good’ can help to maintain news media independence, but research needs to focus on structural inequalities, political alignments and social transformation.

Much more detailed research is needed on the roles of AI systems and news media regulation in encouraging big tech companies, states and other actors to produce and circulate mis- and disinformation.

Little is known about the public’s view of interventions to moderate online mis- and disinformation. There is a slight preference in some studies for individual control over content as compared to platform content moderation or state regulation, but this evidence comes mostly from the United States. Acceptance of strategies to combat mis- and disinformation varies by country, socio-political context, culture and histories of experience with autocratic governments and colonialization.

This area deserves investigation since public acceptance of different methods of moderating mis- and disinformation is likely to influence both their online practices and their trust in news and public institutions.

4.6. TOWARDS DATA JUSTICE IN INFORMATION ECOSYSTEMS

Corporate incentives, strategies and practices involved in controlling data within information ecosystems can lead to epistemic injustice – the privileging of corporate views about how data extractions and monetization should operate, and justifications for the exercise of their monopolistic power (Chapter 8).

The monopolistic power of big tech companies is shown to create harmful discrimination and exclusions in data economies that thrive on data extraction and monetization. The privileging of corporate perspectives can be resisted when an effort is made to reimagine what data justice could be, and to empower individuals and communities to devise proportionate and sustainable uses of data that avoid known biases of business models and of AI systems. The need to design information ecosystems to enable people to express their ideas and identities without experiencing harm is crucial. Research in this area shows that modifying algorithms cannot be expected to address underlying causes of social discord and distrust in society.

¹⁰ RSF. (2024). 2024 World Press Freedom Index – journalism threatened by fake content industry. Reporters Without Borders and UNESCO. (2024, August 14). Statistics on Killed Journalists. UNESCO.

Research is needed on population-level, data-related injustices to reveal how the burdens of datafication are borne disproportionately by certain groups, and how big tech business models lead to biases and exclusions that marginalize populations.

A rights-respecting information ecosystem depends on the capacity for thinking critically about how to govern data, and on recognizing the agency of individuals and groups to resist the power of big tech companies. Data justice initiatives aim to build alternatives to ‘algorithmic injustice’. These enable communities to contest the design of technology systems and the mechanisms for controlling data. Initiatives include digital self-defense tactics, public interest alternative news media, and experiments with community collaborative strategies and municipal efforts to resist ‘smart city’ developments with discriminatory outcomes. They also include the development of community-controlled technologies, proposals for national decentralized data governance frameworks, and work by civil society organizations, researchers and philanthropic organizations aimed at protecting people’s rights.

These initiatives need to be evaluated as there is little systematic research on the efforts underway internationally. A better understanding of these practices is essential to monitor their resource requirements, scalability and capacity to contribute to individuals’ and communities’ sovereignty over the data they provide. Improved insight would help to enable knowledge from diverse sources to inform the future development of information ecosystems, contributing to a paradigm shift that positions the Global Majority World as an equal stakeholder in dialogue about the governance of information ecosystems.

Decolonizing research on data governance and the other issues addressed in this report is essential for Global Majority World experience to inform policy and practice, both within the Global Majority world as well as the Global North.

5 Limitations of this Report

This assessment of research is limited in several important ways (see Chapter 1, Section 5). Our critical review of research was not designed to make specific recommendations to policy makers. An analysis of studies of material socio-economic inequalities in people’s lives is discussed only to the extent that socio-economic conditions are mentioned in research that is cited on other issues; these conditions were not a primary theme. Our principal focus was on the upper service applications layer of information ecosystems, although our discussion of network neutrality issues and zero-rating contracts does touch on issues at the lower infrastructure layer.

We have emphasized the imbalance in Global North and Global Majority World research, and how this favors viewpoints and conclusions of Global North experience throughout this report. Several additional salient fields of research were outside the focus of our work: ‘digital divides’, cybersecurity, securitization, geopolitics and ‘digital sovereignty’, the economic geography of digital labor markets and the (micro)economics of digital markets.

The focus in this report is mainly on country-level experiences and institutions, not on micro-level or sectoral experience. We have not included technology ‘use cases’ or technology application case studies. Finally, our analysis is inevitably limited insofar as all research is guided by the research questions that are posed by research communities, available funding and researchers’ access to data.

6 A Final Word on What Should Be Done

A high concentration of research and research funding in the Global North was confirmed by our critical analysis of state-of-the-art research. This does not help to counter the view that the future of information ecosystems and democracy is one where the Global Majority World emulates best practice in the Global North. Questioning this view is essential, and as above we emphasize the need for work to decolonize research in all fields that inform policy, strategy and practice.

Our analysis has highlighted a major tension. Some research welcomes rapid changes in digital technologies (including AI systems), expecting that harms will, in time, be mitigated. Other research acknowledges the many benefits of new technologies, but emphasizes that their design and use is a result of unequal power relationships that need to be addressed. In this context, the dominant logics of big tech business models, and rules and norms governing information ecosystems, are seen in some of the literature as perpetuating inequalities and injustice. This tension helps to explain why some research emphasizes concepts and responses aimed at risk mitigation as new technologies, such as GenAI, come on the market, while other research emphasizes broader responses to unequal distributions of power, the monopolization of data markets and evidence that this leads to a privileging of economic value over human rights protection.

This report calls attention to the strengths and weaknesses of both research traditions. We emphasize that achieving the *Global Digital Compact's* goal to address technology-facilitated violence, hate speech and mis- and disinformation requires research on the impacts on individuals and on the broader implications of digital technologies, data monetization and monopolistic market structures for democracy.

The 'so what' question was put to us many times during this project – so what can be done now?

Key areas for future research are identified in Sections 3 and 4 above.

This report is based mainly on academic research, but it also benefits from research conducted by other organizations. In particular, research is frequently undertaken or commissioned by **civil society organizations**. These organizations include research think tanks, fact-checking organizations and other non-governmental independent non-profit organizations (in this report, we cite 118 of these organizations – 27% Global Majority, 26% Global coverage).

These organizations are playing a vital role – working with academic researchers – in calling attention to big tech exploitative business practices and proposing remedies such as devising data governance practices for data justice. They are building alternative data governance frameworks aligned with human rights commitments. Their work on local, community and municipal data governance frameworks, and on proposals to introduce decentralized data governance at the national level, is essential to the future health of information ecosystems and to whether democracy flourishes in the future. In addition, civil society organizations are working on monitoring and/or countering the manipulation of information and on media literacy programs.

These activities need improved support in all regions of the world.

Our analysis also points clearly to actions that should inform government policy and private sector practice. We do not make specific recommendations, but offer guidance to public sector and corporate actors (see Chapter 9, Section 6, for an extended list).

Policy makers must take steps to tackle power asymmetries, independently monitor human rights infringements, combat mis- and disinformation, strengthen the transparency and accountability of big tech company strategies and practices, ensure that media and information literacy (MIL) and AI

literacy initiatives are supported and influence research priorities.

Big tech companies must take action to ensure that the public commitments they make to promote safe and democratic online spaces are delivered. This includes changing business strategies to reduce or eliminate the harms associated with data monetization and ensure their practices are aligned with international human rights commitments. Other actions include increasing transparency, and engaging in meaningful consultation with service and technology users and ensuring that content moderation processes are well resourced and accountable.

The map is generated by the OID on the basis of the report's text using GarganText – developed by the CNRS Institute of Complex Systems. Starting from a co-occurrence matrix generated from report's text, GarganText forms a network where words are connected if they are likely to occur together. Clustering is conducted based on the Louvain community detection method, and the visualisation is generated using the Force Atlas 2 algorithm.

xxxii

ACKNOWLEDGMENTS

THE PREFIGURATION GROUP

The International Observatory on Information and Democracy's project was the result of a long reflection about the need to scientifically assess the impact of information ecosystems on democracy and was initiated by the late President of the Forum on Information and Democracy, **Christophe Deloire**. This critical work would never have been possible without his tireless engagement and inspiring leadership. The Forum's team dedicates this report to his legacy, which laid the solid foundations for the Observatory's mission of fostering a more informed and democratic society.

Christophe's ambitious vision translated into a comprehensive prefiguration phase managed by the Forum's then Operations Director, now Executive Director, **Camille Grenier**, who was instrumental in engaging with critically relevant experts to further define the project's ambition.

The Forum's team extends its heartfelt gratitude to the Observatory's prefiguration group, co-chaired by **Shoshana Zuboff**, author of *The Age of Surveillance Capitalism* and Professor Emeritus at the Harvard Business School, and **Ángel Gurría**, former Secretary-General of the OECD. The group played a crucial role during the prefiguration phase, by defining the objectives, methodology and operational framework of the Observatory.

Prefiguration Group Members:

- **Virgilio Almeida**, Professor Emeritus, Universidade Federal de Minas Gerais (Brazil)
- **Jim Balsillie**, Founder of the Center for International Governance Innovation (CIGI) (Canada)
- **Jean-Marie Guéhenno**, diplomat, former United Nations Deputy General Secretary (France)
- **Miguel Poiares Maduro**, Chair, European Digital Media Observatory, European University Institute (Portugal)
- **Burhan Sönmez**, President, PEN International (Turkey/UK)
- **Maria Ressa**, CEO of Rappler, Nobel Peace Prize laureate 2021 (Philippines)

This thorough consultation and collaboration process led to a landmark prefiguration report coordinated by **Chloé Fiodiere** and **Florian Forestier**, who were both instrumental in designing the Observatory's mission and then handing over to the Head of the Observatory, Iris Boyer, who was tasked with setting up the project's governance and implementation roadmap, and transforming words into action.

THE OBSERVATORY TEAM

This report was prepared by a team led by **Professor Emerita Robin Mansell** of the Department of Media and Communications, London School of Economics and Political Science and Scientific Director of the International Observatory on Information and Democracy. The Forum's team would like to express its heartfelt gratitude for her instrumental vision and outstanding leadership throughout the project, and her exceptional investment in time and energy. Her tremendous contribution to this project has been truly humbling and immensely inspiring.

The Forum's team would also like to warmly thank **Professor Rob Procter** of the Department of Computer Science, University of Warwick and Alan Turing Institute for Data Science and AI, UK, for his writing, advisory and editorial support throughout this research work as Lead Rapporteur.

The team is particularly grateful to the Observatory's rapporteurs who synthesized existing evidence and co-wrote the chapters of this report. We thank this international and multidisciplinary team for their constant dedication, insightful perspectives and heartfelt professionalism in the process.

Rapporteurs:

- **Flavia Durach**, PhD, Associate Professor, Head of Department, Department of Digital Communication and Emerging Media, National University of Political Studies and Public Administration, news media rapporteur (Romania)

- **Matthias C. Kettemann**, Professor and Head of Department, Department of Theory and Future of Law, University of Innsbruck; Research Program Leader, Alexander von Humboldt Institute for Internet and Society and Leibniz Institute for Media Research, Hans-Bredow-Institute, AI systems rapporteur (Austria/Germany)
- **Théophile Lenoir**, PhD candidate at the University of Milan and guest doctoral student at the Sciences Po Medialab, news media rapporteur (France)
- **Gyan Prakash Tripathi**, Lawyer and Policy Analyst, data governance rapporteur (India)
- **Emily Tucker**, Executive Director, Center on Privacy & Technology at Georgetown Law, data governance rapporteur (USA)

Strategic overview and project management support was provided by the permanent staff of the Observatory, led by **Iris Boyer**, Head of the Observatory, and coordinated by **Emma Gruden**, Project Officer. The team also extends its immense gratitude to **Michael Bak**, former Executive Director of the Forum, for his leadership – particularly during the critical milestone of the Observatory's launch and early drafting of the report – and to **Alice Boeffard-Dossierre**, **Chiara Gottardo** and **Giovanni Maggi**, who interned at the Observatory. **Katharina Zuegel** provided extensive helpful comments.

THE STEERING COMMITTEE

The Forum's team would like to thank the members of the Observatory's Steering Committee, chaired by **Dr Courtney C. Radsch**, Director of the Center for Journalism and Liberty at the Open Markets Institute, Fellow at UCLA Institute for Technology, Law and Policy, and Fellow at the Center for Democracy and Technology (USA), and co-chaired by **Professor Gustavo Cardoso**, Media Sociologist, Iscte-IUL, and Director of OberCom – Media Observatory (Portugal), for their leadership and guidance throughout the preparation of this report:

- **Luca Belli**, Professor of Digital Governance and Regulation, Fundação Getulio Vargas (FGV) Law School, Rio de Janeiro (Brazil)
- **Niva Elkin-Koren**, Professor at Tel-Aviv University Faculty of Law, and Faculty Director of

the Chief Justice Meir Shamgar Center for Digital Law and Innovation (Israel)

- **Helani Galpaya**, CEO of LIRNEasia (Sri Lanka)
- **Timothy Garton Ash**, Professor of European Studies and Isaiah Berlin Professorial Fellow, St Antony's College, University of Oxford, and Senior Fellow at The Hoover Institution, Stanford University (UK and USA)
- **Frances Haugen**, Co-Founder of Beyond the Screen, algorithmic product expert, and advocate for accountability and transparency in social media (USA)
- **Jeanette Hofmann**, Professor at Freie Universität Berlin, Research and co-founding director of the Alexander von Humboldt Institute for Internet and Society (HIIG) (Germany)
- **Jhalak Kakkar**, Executive Director, Centre for Communication Governance, National Law University Delhi, and Visiting Professor, National Law University Delhi (India)
- **Ansgar Koene**, Global AI Ethics and Regulatory Leader, EY (Belgium)
- **Admire Mare**, Associate Professor and Head of the Department of Communication and Media, University of Johannesburg (South Africa)
- **Nnenna Nwakanma**, digital policy, advocacy and cooperation strategist (Côte d'Ivoire)
- **Pier Luigi Parcu**, Director of the Centre for Media Pluralism and Media Freedom, European University Institute, and Director of the Centre for a Digital Society, European University Institute (Italy)
- **Jeremy Rifkin**, economic and social theorist, bestselling author of 23 books, and President of the TIR Consulting Group LLC (USA)
- **Ghassan Salamé**, former Minister of Culture of Lebanon, and Emeritus Professor of International Relations at Sciences Po Paris (France and Lebanon)
- **Sonja Solomun**, Deputy Director of the Centre for Media, Technology, and Democracy, McGill University (Canada)
- **Nicol Turner Lee**, Senior Fellow in Governance Studies, and Director of the Center for Technology Innovation, Brookings Institution, Washington, DC (USA)
- **Stefaan Verhulst**, Research Professor, New York University (USA), Co-Founder of the Governance Lab, and Co-Founder of Data Tank (Belgium)

- **Natalia Andrea Zuazo**, author of *Guerras de internet* and *Los dueños de internet*, Director of SALTO, and UNESCO consultant (Argentina)

RESEARCH ASSESSMENT PANEL CONTRIBUTORS

Members of the extended team included expert contributors to the Research Assessment Panels (RAPs), who were instrumental in aggregating and synthesizing sources and commenting on the report. The Forum's team and the Scientific Director express their sincere gratitude for the RAP contributors' time, continuous engagement and support:

- **AI RAP:** Yik Chan Chin (China/UK), Giovanni De Gregorio (Portugal), Paula Gori (Italy), Argyro Karanasiou (UK), Shyam Krishna (UK), Ioannis Kompatsiaris (Greece), Galen Lamphere-Englund (UK), Monica Lopez (USA), Devika Mehta (India), Vincent Obia (UK), Immaculate Odwera (USA), Umut Pajaro Velasquez (Colombia), Anass Sedrati (Sweden), Theodora Skeadas (USA), Sharon Strover (USA), Scott Timcke (South Africa), Evelyne Tauchnitz (Switzerland), Shenja van der Graaf (the Netherlands).
- **Media RAP:** Anurag (Canada), Jens Barland (Norway), Jose Luis Benitez (El Salvador/Honduras), Mădălina Boțan (Romania), Paula Gori (Italy), Minna Horowitz (Finland), Ioannis Kompatsiaris (Greece), Jun Liu (Denmark), Beatriz Lopes Buarque (UK), Reyhana Masters and Stephen D. McDowell (USA), Sacha Meuter (Switzerland), Katie Pentney (UK), João Alexandre Peschanski (Brazil), Pauline Renaud (UK), Rose Marie Santini (Brazil), Theresa Josephine Seipp (the Netherlands), Marlyn Tadros (Egypt/USA) and Kate Wright (UK).
- **Data governance RAP:** Joren Baillière (Belgium), Poncelet Ileleji (The Gambia), Monica Lopez (USA), Jenna Manhau Fung (Hong Kong SAR), Siddhartha Menon (USA), Umut Pajaro Velasquez (Colombia), Sharon Strover (USA), Michaela Sullivan-Paul (the Netherlands), Sahaj Vaidya (USA).

We are especially grateful to numerous Steering Committee and Research Assessment Panel (RAP) members who took the time to comment

in detail on earlier drafts of the report and shared constructive feedback during our meetings.

Within the Steering Committee, special thanks go to Luca Belli, Helani Galpaya, Jeanette Hofmann, Ansgar Koene, Nnenna Nwakanma, Pier Luigi Parcu, Courtney Radsch and Natalia Zuazo for their sustained engagement and critical feedback throughout the editorial process.

From the RAPs, we received very helpful comments and edit suggestions from Mina Aslam, Jens Berland, Beatriz Burarque, Jenny Fung, Giovanni De Gregorio, Poncelet Ileleji, Devika Mehta, Siddhartha Menon, Sasha Meuter, Vincent Obia, Joao Peshanki, Pauline Renaud, Nikos Sarris, Anass Sedrati, Theresa Seip, Sharon Strover, Michela Sullivan-Paul and Kate Wright.

In a few instances, textual contributions were offered and incorporated into the report (Anurag, Jeanette Hofmann, Sharon Strover).

EXTERNAL CONTRIBUTORS

The Forum's team warmly thank the hundreds of academics, members of civil society organizations and think tank representatives who took part in the **Stakeholder Advisory Group** consultation meetings on 4 September 2023 and 27 June 2024, and who shared their insights and perspectives during the production of the report. Likewise, the team thanks the **representatives of States and public institutions** of the Partnership for Information and Democracy for their participation in the consultation meetings held on 7 September 2023 and 25 June 2024.

The team also benefited from the input of **external contributors** who shared their research through the **global calls for contributions**: Lauro Accioly Filho (Brazil), Theodora Dame Adjinn-Tettey (South Africa/Ghana), Susan Ariel Aaronson (USA), Muhammed Alakitan (UK), Gene Allen (Canada), Tsiry Rambel Andrianisa (Madagascar), Olivier Arifon (Belgium), Homa Azodi (Canada), Brian Ball (UK), Habmo Birwe (Senegal), Anita Breuer (Germany), Emma Briant (Australia), Bastien Carniel (France), Kuan-Wei Chen (Taiwan), Lyantoniette Chua (Philippines), Joel Christoph (Italy), Julian D. Cortes (Colombia),

Tim Davies (USA), Laurence Dierickx (Norway), Astou Diouf (Senegal), Maira Elahi (Canada), Oscar Espiritusanto (Spain), Jean-Louis Kedieng Ebongue Fendji (Cameroon), Adalberto Fernandes (Portugal), Agustin Ferrari Braun (the Netherlands), Giovana Fleck (the Netherlands), Terry Flew (Australia), Elaine Ford (Peru), Margaret Gallagher (Ireland), Diego Garcia Ramirez (Colombia), Anna Gibson (USA), Rachel Griffin (France), Armando Guio (Colombia), Marci Harris (USA), Sarah Hayes (UK), Petros Iosifidis (UK), Dimitrios Kalogeropoulos (UK), Irini Katsirea (UK), Anselm Küsters (Germany), Aurora Labio-Bernal (Spain), Horacio Larreguy (Mexico), Michel Leroy (France), DeVan Hankerson Madrigal (USA), Sabeedah Mahomed (UK), Fernando Trincado Moraes (Brazil), Judith Murungi (Uganda), Waqas Naeem (Pakistan), Gabriel Nicholas (USA), Alessandro Paciaroni (Estonia), Charis Papaevangelou (The Netherlands), Stella Meyer (Belgium), Henna Paakki (Finland), Ryan Payne (Australia), Bethia Pearson (UK), Phil Ramsey (Northern Ireland), Pradipa P. Rasidi (Indonesia), Anna Romandash (Ukraine), Felix M. Simon (UK), Anastasia Stasenko (France), Edward Steinmueller (UK), Zea Szebeni (Finland), Melissa Terras (Scotland), Jorge Vázquez-Herrero (Spain), Karen Vergara Sánchez (Chile), Andrew White (UK), Dolly Wong (Indonesia), Gaston Wright (Argentina), Paul-Alain Zibi Fama (Cameroon).

Special thanks are extended to several **experts** who were interviewed for this report:

- J. Khadijah Abdurahman, Editor-in-chief, Logic(s), USA
- Natalia Aruguete, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Argentina
- Jhalak Kakkar, Executive Director, Centre for Communication Governance, National Law University Delhi, and Member of the Observatory's Steering Committee
- Eugenia Mitchelstein, Associate Professor of Communication, Universidad de San Andrés, Argentina
- Syed Nazakat, CEO of DataLEADS, India
- Rasmus Nielsen, Director of the Reuters Institute for the Study of Journalism, Professor of Political Communication, University of Oxford, UK during the project, now Professor of Communication, University of Copenhagen, Denmark.

- Marianna Poyares, PhD candidate in Philosophy, The New School for Social Research, USA.

This project also benefited from inputs from **think tanks, research institutes** and **other organizations** including the Institute for Strategic Dialogue (ISD), the Internet Freedom Foundation, IT for Change and UNESCO, as well as national regulators, with valuable insights provided by Ofcom's Jordan Ogg and Mark Pearson, observers during the RAP meetings.

The Observatory is grateful to all those who gave us permission to use their tables and figures.

EDITING, TRANSLATION, DESIGN AND LAYOUT

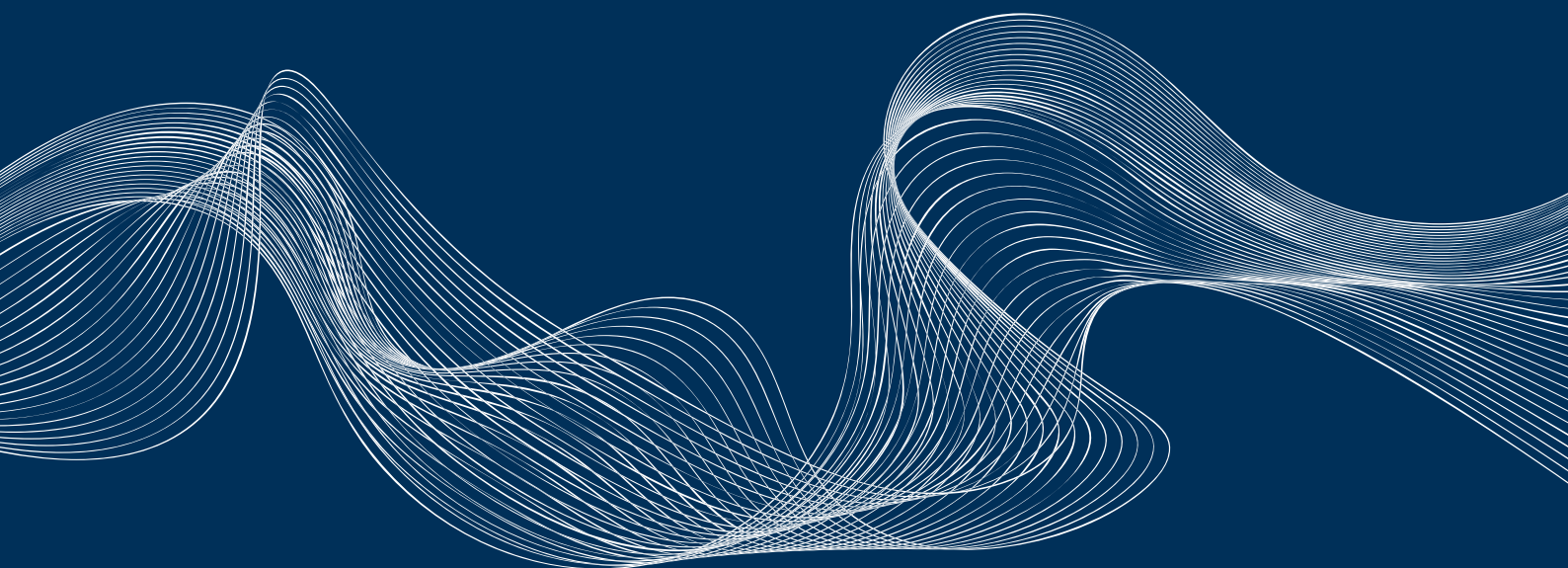
The report benefited from Dawn Rushen's excellent copyediting skills. It was designed by Matthieu Mercier, and the Executive Summary was translated by Nathalie Reis.

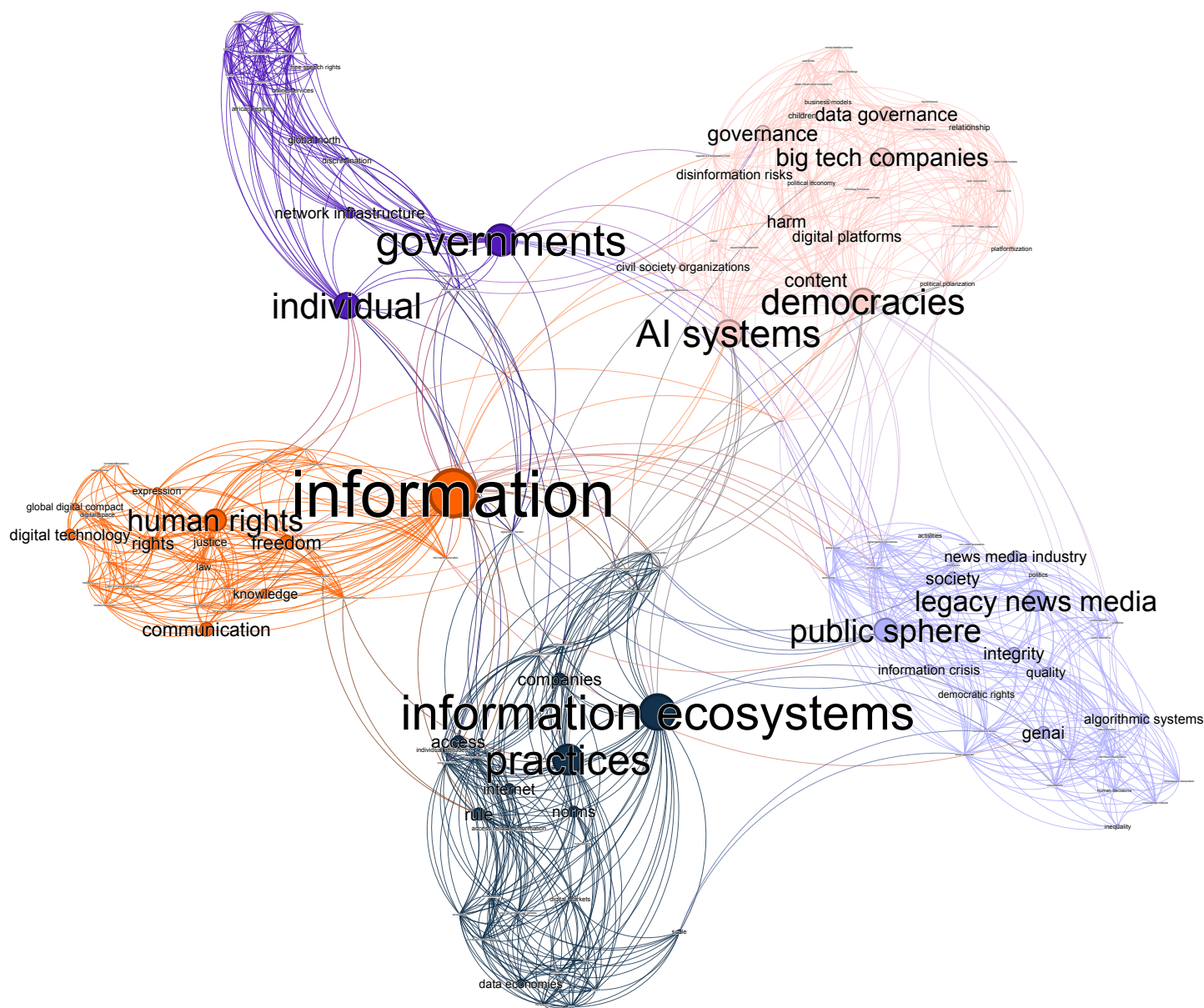
FINANCIAL SUPPORT

Last, but not least, the Forum's team thanks its donors for their trust and support throughout this process. The Observatory project was funded by the Postcode Foundation, the Ford Foundation, Luminate and the French Ministry of Foreign Affairs.

CHAPTER 1

INFORMATION ECOSYSTEMS AND DEMOCRACY





This map represents a statistical summary of the thematic content of this chapter. The network graph represents relations between the words in the chapter, placing them closer to each other the more they are related. The bigger the node, the more present the word is, signalling its role in defining what the report is about. The colors represent words that are closely related to each other and can be interpreted as a topic.

The map is generated by the OID on the basis of the chapter's text using GarganText – developed by the CNRS Institute of Complex Systems. Starting from a co-occurrence matrix generated from chapter's text, GarganText forms a network where words are connected if they are likely to occur together. Clustering is conducted based on the Louvain community detection method, and the visualization is generated using the Force Atlas 2 algorithm.

[Link to the interactive map here](#)

*‘The right to know is the right to live’
(Aruna Roy).¹*

1 Introduction

The United Nations’ *Global Digital Compact* asserts that ‘digital technologies are dramatically transforming our world’.² Agreed in September 2024, the text insists that human oversight of technology is needed to identify and mitigate the risks for humanity.³ In emphasizing technology-driven transformation and human choice relating to risk mitigation, it is easy to lose sight of the fact that the design and development and the beneficial and harmful uses of technology are not dictated by technology; rather, they are the result of human decisions and action. What technology designers, corporate, government and individual decision-makers believe is appropriate technological development is not immutable: transformation depends on power relationships in societies, the presence and strength of countervailing forces, and ‘whether those who are not in the corridors of power can organize and have their voices heard’.⁴

This report is a critical analysis of research in the Global North and the Global Majority World⁵ that informs us about the interdependent relationships between the cultural, social, political, economic and technological components of information ecosystems.⁶ It focuses on what interdependence means for the integrity of information and for informed

democratic participation in the public sphere. It means understanding questions about the generation and circulation of mis- and disinformation as symptoms of broader and complex changes in society and as important amplifiers of these changes.⁷ The report investigates how these reciprocal relationships are playing out in the news media industry, in the development and use of artificial intelligence (AI) systems,⁸ and in the ways that data is generated, processed and controlled.

Information ecosystems are implicated in the integrity of information (the quality of public discourse), the fairness of political processes, the protection of media freedoms and the resilience of public institutions.⁹ The report addresses three thematic areas with a cross-cutting theme of mis- and disinformation: **media, politics and trust; artificial intelligence, information ecosystems and democracy; and data governance and democracy**. The analysis is based on a large collection of research sources (3,095 of which 1,664 are cited) including academic publications, reports and other materials.¹⁰ Insight into whether changes in these areas are consistent with the protection of fundamental human rights is of special interest when democracy is troubled – not only by changes in information ecosystems, but also by multiple instances of injustice.

¹ Roy, A. (2004, n.p.), Indian social activist, professor, union organizer and former civil servant and President, National Federation of Indian Women.

² UN (2024b, paras 1, 7). The Compact sets out five objectives: (1) close all digital divides and accelerate progress across the Sustainable Development Goals; (2) expand inclusion in and benefits from the digital economy for all; (3) foster an inclusive, open, safe and secure digital space that respects, protects and promotes human rights; (4) advance responsible, equitable and interoperable data governance approaches; and (5) enhance international governance of artificial intelligence for the benefit of humanity. Its ambitions are discussed in Section 4 of this chapter. See also UN (2024d).

³ UN (2024b, para. 3).

⁴ Acemoglu & Johnson (2023, p. 29).

⁵ See Campbell-Stephens (2021) for a discussion of naming practices. It is difficult to settle on a set of definitions for groups of people or countries. We use ‘Global Majority World’ (and not without criticism) as a collective reference to 85% of the world’s population who live in low- and middle-income countries and who are of Indigenous, African, Asian or Latin American descent (sometimes extended to people of dual heritage, and minority ethnic groups who are racialized within countries and not classed as ‘White’). ‘Global North’ is used to refer to those not included in the Global Majority World, generally from the wealthy, industrialized countries. Regions and countries or specific groups are discussed as appropriate. See Anthony *et al.* (2024); Lawrence (2022); Patrick & Huggins (2023).

⁶ The terminology used in this report is discussed further in Section 3 of this chapter and in Appendix: Methodology.

⁷ See Tay *et al.* (2024, p. 1), who note that misinformation ‘depending on individual and contextual factors ... can be both a symptom and a cause’, and that multidimensionality is important. Our socio-technical perspective directs attention to reciprocal relationships between components of complex societal systems, combined with a political economy perspective that directs attention to power structures and relationships.

⁸ Mueller (2024, p. 2) argues that the label ‘AI’ is unhelpful, since what we are discussing is ‘digital ecosystems’ including ‘computing devices, digital networks, digitized data, and software programs’. He argues that references to ‘AI governance’ are becoming meaningless. We refer to AI systems and to specific components of AI systems whenever possible (the rationale is explained in Section 1, Chapter 3).

⁹ The main focus is on the resilience of public institutions and criticisms of those institutions when they are complicit in injustice. The resilience of individuals is discussed in relation to interactions with online content (across the chapters) and self-defense measures (in Chapter 8), but the aim is not to position individuals themselves as ultimately responsible for accommodating injustices arising from datafication processes. See Banaji (2024).

¹⁰ Of 1,664 cited sources, 65.5% classified as Global North, 22.5% Global Majority World, and 12% Global. See Appendix: Methodology.

This assessment of research is not about identifying technology or mis- and disinformation as the *principal* cause of democratic fragility. Rather, while much research is mainly concerned with the impacts of digital technology or mis- and disinformation on society or individuals, we aim to assess research findings in the context of how and why information or technology is problematic, for whom, and what is being done, or could be done, to mitigate problems. Strengths, weaknesses and gaps in research are identified to improve understanding of how democratic decision-making and justice might be achieved in data-intensive economies.

Section 2 of this chapter explains the interdependence of datafication processes and democracy, highlighting why mis- and disinformation has become a prominent focus of research. Section 3 introduces what we understand to be the principal components of an information ecosystem, and explains key concepts used in this report, including the public sphere and the international human rights commitments that are central to any assessment of information and communication. Section 4 explains how mis- and disinformation are understood in much of the policy literature, and why the norms and rules for governing the production, circulation and use of data and information are crucial issues, especially at a time when democracy itself is troubled in many countries around the world. Section 5 then explains the structure of this report, and outlines the content of the chapters that follow.

2 Setting the Context: Datafication and Democracy

Policy in multiple countries is saturated with claims about the harms of online mis- and disinformation. Warnings about an 'information crisis' are galvanizing governments, companies and civil society organizations to develop principles, guidelines and strategies for combating mis- and disinformation. The World Economic Forum's *Global Risks Report 2024* put mis- and disinformation risks at the very top of its list of perceived threats.¹¹ Principles are being agreed internationally for the responsible development and use of AI systems.¹² UNESCO, for example, has produced a set of guidelines for governing digital platforms.¹³ Together with the *Global Digital Compact*, these aim to ensure that those who design, operate or participate in information ecosystems (including network infrastructures, data and content) adhere to international human rights commitments.¹⁴

A critical multidisciplinary assessment of research on the interdependence of information ecosystems, the public sphere and democracy is crucial in view of evidence that democracy is troubled by changes in information ecosystems around the world. For example, internet freedom declined globally for the 14th year in a row in 2024. In three-quarters of the 72 countries examined by the Freedom House *Freedom on the Net* report, online users were arrested for non-violent expression and people were physically attacked or killed for their online activities in at least 43 countries.¹⁵ In addition, since 1993, 1,701 journalists have been killed, with 50% of these deaths occurring outside conflict zones.¹⁶

¹¹ WEF (2024).

¹² For example, In September 2024 at the United Nations level, the Pact for the Future, Global Digital Compact, and Declaration on Future Generations (UN, 2024b) and the Governing AI for Humanity report, which calls for 'a collaborative and learning mindset, multi-stakeholder engagement and broad-based public engagement', and acknowledges that 'whole parts of the world have been left out of international AI governance conversations' (UN, 2024a, pp. 78, 8). See also earlier statements from the G7 (2023); OECD (2022c); UK DSIT (2023).

¹³ UNESCO (2023b).

¹⁴ Including the Universal Declaration of Human Rights (UDHR) (UN, 1948) and the International Covenant on Civil and Political Rights (ICCPR) (UN, 1966).

¹⁵ Freedom House (2024).

¹⁶ UNESCO (2024).

Big tech company decisions influence operations and editorial choices in the news media industry, which is essential for the democratic ordering of society (our first thematic area: **media, politics and trust**). In principle, the news media industry can hold the powerful to account and facilitate the free exchange of accurate information. However, if information is wrong or inaccurate and circulates virally, the quality of public debate suffers. The commercial imperatives of algorithmic-driven and opaque advertising markets mean that political news often appears next to sensationalist content. With legacy media facing competition from podcasts and individual bloggers, the combination of and concentration in the legacy media industry, and the proliferation of online information flows, is creating a financially unstable environment for the gathering and reporting of news.

News media professionals feel pressured to make their content more attention-grabbing to adapt to digital platform affordances, sometimes sacrificing content quality.¹⁷ Concern about what is real and what is ‘fake’ online news is reported to have risen to 59% globally: in the United States to 72% and in South Africa to 81%, both countries that held elections in 2024.¹⁸ Declining trust in the news varies by country, but is concerning. In a global survey in 2024, respondents were asked whether they trusted the news most of the time. Finland recorded the highest overall trust, at 69%, the United States, 32%, France, 31%, Argentina, 30%, Greece, 23% and Hungary, 23%.¹⁹ The contribution of news producers to the public sphere and to whether news media organizations are trusted depends on the context in which they operate – democratic or autocratic – the legal authority under which they operate, and whether pluralism and diversity are encouraged.²⁰ Questions about how the independence and financial viability of news organizations can be sustained are common across countries. The varied responses have

substantial consequences for the independence of the news media and for the protection of democratic rights and freedoms.²¹

Mis- and disinformation circulating at scale is seen as diminishing the quality of the news media and public discourse.²² However, research on the information crisis often neglects the role of legacy news media and the history of propaganda. Analysis focuses principally on the impacts of technological change, neglecting non-technical factors that influence information integrity.²³ In 2024 more than 80 countries and some 3 billion people were set to vote in regional or national elections.²⁴ Wars were being waged in Somalia, between Russia and Ukraine, and involving Israel and other territories and states. In the context of microtargeting, the use of biased AI systems, the rise of ‘deep fakes’, the escalation of cyberattacks and the weaponization of information, there is good reason to be concerned about the integrity of information and the problems faced by the news industry.²⁵

The release to the public of generative artificial intelligence (GenAI) in 2022 means that the tools for information manipulation have become more available and less costly. Algorithmic systems, including large language models (LLMs), contain unavoidable biases that stem from data generation and collection processes that are subject to human decisions (our second thematic area: **artificial intelligence, information ecosystems and democracy**). In addition, given that people’s own biases influence their online behavior and interpretations and uses of information, these are exacerbated when outputs are used to confirm preconceived notions and when AI models are trained on these outputs. In turn, this diminishes the quality of information over time,²⁶ leading to decisions that perpetuate inequality and increase

¹⁷ Chadwick (2017).

¹⁸ Newman *et al.* (2024).

¹⁹ Newman *et al.* (2024).

²⁰ Hallin & Mancini (2004, 2012); Neff & Pickard (2024).

²¹ Altay *et al.* (2023a); Baines & Elliott (2020); Epstein (2020); Fallis (2015); François (2019); Kapantai *et al.* (2021); Milton & Mano (2022, pp. 34, 49); Ó Fathaigh *et al.* (2021); OH-CHR (2021, paras 9–15); Pielemeier (2020); Willems (2014a). For a discussion on how human rights and democracy can be united in governance structures, see Besson (2011).

²² Bennett & Kneuer (2023); Jungheer & Schroeder (2021); Schlesinger (2020); Wasserman (2020a).

²³ Benkler *et al.* (2018); Bolin & Kunelius (2023); Hyzen (2023); Tsfati *et al.* (2020).

²⁴ Harbath (2023) discusses the difficulties of counting elections.

²⁵ Craig *et al.* (2023); Holt (2023); O’Connor (2022), all Institute for Strategic Dialogue (ISD), an independent organization; see also Briant (2024); Caulfield *et al.* (2023); Forum on Information and Democracy (2023, 2024a).

²⁶ See Dolata *et al.* (2022); Kop (2020), and Corbett-Davies *et al.* (2017) supported in part by the John S. and James L. Knight Foundation and Hellman Fellows Fund, US.

vulnerabilities.²⁷ These developments can negatively influence the way citizens understand themselves as political actors, with disproportionately negative effects on marginalized people.²⁸

The big tech companies use their algorithmic systems to analyze behavior and keep people engaged in interactions that generate data.²⁹ ‘Datafication’ enables companies to transform everyday actions into quantified data that is used for real-time tracking and predictive analysis.³⁰ The use of computational methods results in the untransparent manipulation of information and communication flows.³¹ Thus, this ‘platformization’ of information means that the economy and multiple spheres of public and private life are influenced by the choices of these companies.³² Although the big tech companies dominate in providing an infrastructure for information ecosystems, other digital intermediaries play an important role, for example network operators and web-hosting companies. These have the capacity to alter information ecosystems, for example by shutting down the internet or taking websites offline unilaterally or under pressure from governments.³³ The big tech companies and states aiming to be leaders in the global economy argue that their competitiveness and national (regional) economic growth depend on greater efficiencies in the collection and monetization of data, and they claim that ‘technological accelerationism’ is good for humanity.³⁴ They insist that online interaction generates ‘raw’ or ‘neutral’ data that belongs to no one (until it is appropriated by them).

This is the context in which an information crisis has come to the top of the policy agenda. The technology companies’ business models and practices are implicated in what critical scholarship refers to as ‘surveillance capitalism’ or ‘data colonialism’.³⁵ Individuals and societies are being comprehensively surveilled for data extraction and products, and the monopolization of information generates revenue by converting data (including public data) into private information assets (our third theme: **data governance and democracy**). Digital platforms and AI systems are opening a space for a more reciprocal dynamic between humans and technology that is potentially beneficial. However, how these components of information ecosystems operate is decided largely by these companies within the legal frameworks that are put in place by governments. These governance arrangements determine what information ‘can appear, how it is organized, how it is monetized, what can be removed and why, and what the technical architecture allows and prohibits’.³⁶ When, for instance, someone shows interest in a type of political content, an algorithm is likely to overemphasize similar viewpoints in their feed, narrowing the range of information they see. It is broadly accepted that these practices, combined with the positioning of individual freedom as the enemy of equality and solidarity, are implicated in social and political instability.³⁷ The datafication practices do not fully explain political or economic divisions in society – these ‘exist before and beyond’ these companies’ activities,³⁸ and online mis- and disinformation are not the only contributing factors.³⁹ However, when information is

²⁷ See, for example, Wang *et al.* (2024), on issues of vulnerability in the use of AI systems.

²⁸ Horowitz *et al.* (2024); Liveriero (2020, p. 787). Epistemic rights refer to the requirement that to achieve equality in decision-making, it must be guaranteed that truthful information and knowledge are available to all. ‘Epistemic rights are about knowledge – not only about being informed, but also about being informed truthfully, understanding the relevance of information, and acting on its basis for the benefit of oneself and society as a whole’ (Nieminen, 2024, p. 15).

²⁹ Nieborg & Poell (2018); Plantin *et al.* (2018); van Dijck *et al.* (2018a). The metaphor ‘platform’ has been criticized for giving a misleading indication of the specific transformation processes (Gillespie, 2010), although it is still used widely in the literature.

³⁰ Transforming offline action into online quantified data enabling tracking and predictive analysis; see Mayer-Schönberger & Cukier (2013).

³¹ Gitelman (2013, p. 7).

³² Poell *et al.* (2019, p. 1).

³³ See Bradshaw & DeNardis (2022) on infrastructure and disinformation; see also Bradshaw *et al.* (2021), supported by the European Research Council (ERC), Adessium Foundation, Civitates Initiative, Ford Foundation, Hewlett Foundation, Luminare, Newmark Philanthropies and Open Society Foundations.

³⁴ Caballero & Monje (2024).

³⁵ See Bennett & Livingston (2023); Couldry & Mejias (2019); Fendji (2024); Lee & Valenzuela (2024); Lehdonvirta (2022); Mejias & Couldry (2024); Trappel (2019); van Dijck *et al.* (2018a); Zuboff (2019). This work builds on several decades of using untransparent offline advertising techniques and now online personalization systems; see (McGuigan, 2023).

³⁶ Gillespie (2010, p. 359).

³⁷ Calhoun *et al.* (2022).

³⁸ Tonnies (1957, p. 140), first published as *Gemeinschaft und Gesellschaft* in 1887.

³⁹ Aruguete & Calvo (2023); Zuazo & Aruguete (2021).

inaccurate, harmful or illegal (e.g., violent, associated with nationalism, ethno-religious bigotry or misogyny), the risks to individuals and groups can multiply, especially if people cannot discriminate between accurate and inaccurate information.

With the big tech companies turning a blind eye to how they facilitate the generation and circulation of mis- and disinformation, there are multiple efforts to introduce further governance arrangements to force, or to seek a renewal and extension of, voluntary compliance in the responsible management of online services. Companies are marketing advanced digital technologies, including GenAI, as quickly as they can claim to adhere to safety standards and responsible innovation practices. The pace of these developments can sideline or overcome countervailing power mounted through regulatory or civil society action by creating internal information ecosystems governed by private rule-making embedded in automated technologies. Doing so preempts meaningful political deliberation about rights of data ownership, what role data should have in the economy and public sector, how it should inform bureaucracy, and in what contexts data production should be minimized or prohibited. That is, it contributes to the declining health of information ecosystems.

Policy makers in the Global North are developing governance frameworks with the aim of balancing national (or regional) races to achieve leadership in digital markets with commitments to securing the rights of publics by setting norms and rules aimed at improved accountability and transparency of the big tech companies. China, the European Union and the United States, for example, are putting governance arrangements in place to maximize the scale and scope of their data economies while also claiming to

balance respect for international human rights law. Their approaches differ, and we need to understand better which publics and whose interests are being protected.⁴⁰

In the Global Majority World, policy makers often confront decisions taken in the Global North and struggle to govern their information ecosystems in ways that reflect their interests. With the big tech companies monopolizing digital service and data markets, the space for imagining and experimenting with alternatives is diminishing. Other countries and regions see the rule-setting big tech companies as ‘behemoths’, and experience a form of ‘digital imperialism’.⁴¹ This is particularly so when the export of governance models by the big tech companies and governments is couched in the language of aid, cooperation and trade. In the case of the African Union and African countries, research indicates that ‘instruments tend to emulate best practices from other regimes’ with unintended consequences when they not ‘suitable for, or overlook African realities’.⁴²

This report focuses mainly on the experiences of those who are connected to the internet. In 2024 there were 5.4 billion individual internet users – 67% of the world’s population – each of whom can be a social media viewer and potentially a ‘speaker’ if they have affordable access. There were an estimated 5 billion active social media user identities in 2024 (62.3% of the world’s population, not necessarily unique individuals).⁴³ Some 2.6 billion people are not connected, and world connectivity averages tell us little about how people experience their online activity since they hide large disparities: in low-income countries, 20.9% of people use the internet; in high-income countries, the figure is 90.5%.⁴⁴ Social, political, cultural and economic factors also influence the production

⁴⁰ AI Now Institute (2024, p. 19).

⁴¹ See Aaronson & Leblond (2018) and Chen & Gao (2022), supported by the National Social Science Fund of China (NSSFC). Examples of data localization initiatives are India’s Digital Personal Data Protection Act (Government of India, 2023) and the African Union Convention on Cyber Security and Personal Data Protection (African Union, 2014). The latter promotes a unified, continent-wide approach to cybersecurity and data privacy that is said to diverge from the European Union’s General Data Protection Regulation (GDPR), and represents 55 African states. Distinctive strategies are discussed in Duncan (2023) and Andere & Kathure (2024).

⁴² Musoni *et al.* (2024, p. 15) supported by the European Commission.

⁴³ Thompson & Kemp (2024).

⁴⁴ ITU (2024b). In the Global Majority World, fixed internet connectivity is either absent or unaffordable for many, which is partly compensated for by mobile internet connectivity, which can be unreliable; see ITU (2024a, p. 13). And there are large differences – mobile broadband subscriptions per 100 inhabitants in 2024: world average 90.7, low-income country 35.3, high-income country 123.4. The cost of a mobile data and voice high consumption basket (140 min., 70 SMS, 1.5GB, 3G and above) as a percentage of Gross National Income per capita shows big disparities: world average 4.7%, low-income country 18%, high-income country 0.9%. The percentage of people who own a mobile phone: world average 82.8%, low-income country 50.4%, high-income country 95.1% (165 countries). Disparities within countries, especially rural and urban, are just as important as those between countries and regions (Strover *et al.*, 2024).

and circulation of mis- and disinformation when connectivity is achieved,⁴⁵ and information also reaches those who are unconnected, for example people share accounts, communicate in offline social networks and engage with legacy types of news. Especially in the Global Majority World, the economically disadvantaged are offered online service contracts that limit exposure to diverse sources of information. In some countries they are legally obliged to participate in political processes, even though their ability to access reliable information depends on the quality of their access to the internet, and social practices and economic policies that influence that access and political discourse. It is therefore essential to take these factors into account rather than focusing principally on the effects of mis- and disinformation on individual political attitudes and behaviors.

3 Positioning the Research Assessment: Concepts and Definitions

A critical analysis of existing research requires decisions about the terminology and concepts to use in conducting the research assessment. A vocabulary is needed to name the objects and processes that are the focus of a research assessment, and the naming itself is controversial. (We explain the choices of terminology and concepts used in this report in Appendix: Methodology.)

We focus on ‘information ecosystems’ that are comprised of social and material components. Specifically, we define an ecosystem as a *system of people, practices, values, and technologies in a particular environment*, embedding the public sphere within two layers of the ecosystem: a network infrastructure (hardware and software) layer and a service applications layer.⁴⁶ By network infrastructure we mean the hardware and software that supports communication, the standards and protocols, and also the actors that produce the technologies and their values and practices.⁴⁷ By service applications layer we mean the variety of services available to users and the values and practices of those who design and operate services.⁴⁸ Figure 1.1 shows the layers of these information ecosystems – infrastructure layer and services applications layer.

For the service applications layer we focus on the news media industry, which depends on services on the applications layer, the development of AI systems and on the norms and practices that govern how data is produced, processed and used. The corporate sector plays a major role in deciding how the layers are designed and operated, but other ownership alternatives are also of interest.⁴⁹ Information ecosystems are assumed to be in constant flux, and power relations and asymmetries mean that we do not expect these systems to achieve a timeless balance among competing interests; struggles among interested parties are understood to be ongoing.

⁴⁵ In this report, the large body of research on digital divides is not reviewed. Zero-rating policies for internet access are discussed in Section 4.1, Chapter 6, and how socio-economic inequalities and marginalized groups are affected by exclusions and harms on the infrastructure and service applications layers of information ecosystems is discussed in Chapter 8.

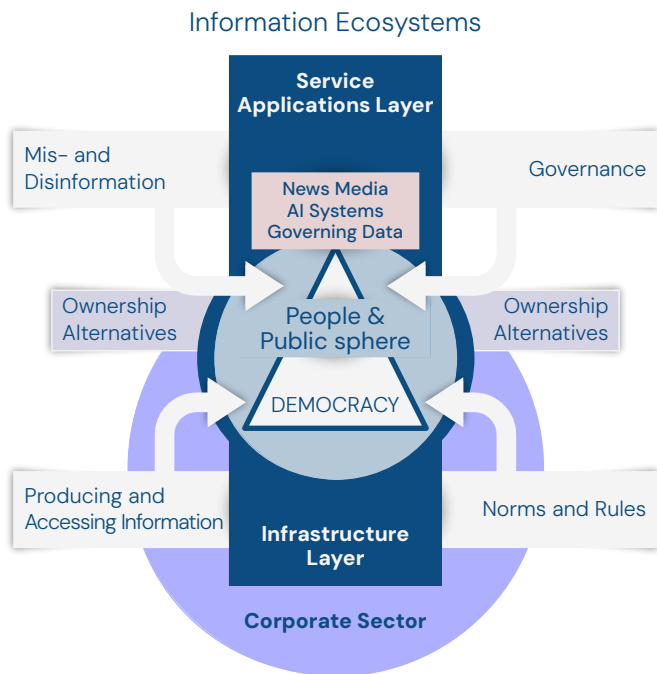
⁴⁶ Modified from Nardi & O’Day (1999, p. 49).

⁴⁷ This would include, for example, cables, data centres, semiconductors and internet access points.

⁴⁸ This would include, for example, cloud services, payment services, search engines, messaging services, app stores, social networking and e-commerce retailers.

⁴⁹ Non-corporate ownership alternatives are discussed in Chapter 8.

Figure 1.1
Information ecosystems and the public sphere



Source: Authors of the report

‘Information ecosystems’ terminology is present in policy debates and in some of the academic literature. Since this report aims to provide a resource for academics, researchers working with civil society organizations and policy makers and regulators, we chose to use this concept.⁵⁰ However, as shown in the center of Figure 1.1, the key emphasis of this research assessment is on how people and their communities interact with information ecosystems, and how this influences the ‘public sphere’, and democracy.⁵¹ We recognize that interactions within digitalized information ecosystems and attempts to have influence in the public sphere for many occur ‘during ongoing deprivation, campaigns of disinformation, police brutality and/or military atrocity’.⁵²

Dependencies and interactions in information ecosystems set the conditions for whether a public sphere can flourish. The public sphere concept is often associated with Eurocentric assumptions about liberalism and an idealized conception of rational communicative action. However, it does sensitize us to how the public sphere became dominated by private interests. An historical perspective helps to shed light on contemporary inclusions and exclusions, and the need to be open to the potential for ‘new forms of solidarity, new forms of intimacy, new forms of collective action, and new forms of identification – in short, new forms of being “public” with strangers’.⁵³

Information ecosystem interactions occur at local, national, regional and global levels and are informed by distinctive and complex social, cultural, political and economic conditions (not shown in Figure 1.1). When the interactions of these components facilitate mis- and disinformation, they can be an impediment to democratic flourishing, and this is treated as a values-based judgment. Thus, claims in the research literature about the ‘health’ or otherwise of information ecosystems are treated as values-based judgments, not simply as indicative of the need to eliminate a pathogen from the system.

There are many approaches to the definition of information ecosystems. Another is a ‘rhizomatic’ systems approach’ which embraces human and non-human elements, defining a healthy information ecosystem as:

A balanced and well-functioning system of information creation, exchange, flow and utilization. It is characterized by the presence of diverse and pluralistic sources of information, information integrity; responsible information production, management and

⁵⁰ See Appendix: Methodology for details on the use of the term ‘ecosystem’ and the health of an information ecosystem. The use of the ‘ecosystem’ concept in this report is guided mainly by socio-technical and political economy theories. An effort is made to draw distinctions between ideal (normative) systems and values and individual and institutional practice. We are concerned with power relationships and struggles among actors over the design and operation of principally the applications layer of information ecosystems. The exercise of individual and collective power is understood to involve agency, to be values-based and to be operating at both the individual and institutional level. See Jasanoff (2015); Mansell (2012); Suchman (2023) for discussions of socio-technical and political economy traditions; see also Musiani (2022), funded by Agence nationale de la recherche (ANR); We are also informed by one branch of systems theory to explain the dynamics of changes in information ecosystems, see Radsch (2023e, p. 1) where the focus is on networks of humans and non-humans and not on the individual information consumer.

⁵¹ See Appendix: Methodology for a discussion of the ‘public sphere’ concept. The coexistence of multiple public spheres where people participate in public life with unequal power is acknowledged (Fraser, 1992), especially in the Global Majority World, where those on the margins are affected by colonialism (Dutta & Pal, 2020). See Cammaerts (2024, p. 27); Ehrenfeld (2020, p. 308); Habermas (2015), first published in English in 1989, in German in 1962; Habermas (2022); Štětka & Mihelj (2024b). For criticism, see Banaji (2024); de Sousa Santos (2018); Splichal (2022b, p. 213).

⁵² Banaji (2024, p. 13).

⁵³ Ehrenfeld (2020, p. 308).

securitization practices; and the ability of individuals and communities to effectively access, analyze, and use information for decision-making, culture-creating, community-building, and accountability'.⁵⁴

This approach is similarly concerned with system interconnectedness and dependencies. It conceives of an information ecosystem that is organized in non-hierarchical and non-linear ways and where there is no 'dominant power controlling the flow of information'.⁵⁵ Instead of putting individuals, citizens or community at the center of the analysis of ecosystem changes, it centers information, technologies, institutions, norms and practices.

The approach in this report seeks to encompass people's and their communities' engagement in the public sphere which is enabled or disabled by information ecosystems and institutionalized norms and practices associated with information generation, distribution and consumption as well as technologies.

The components of information ecosystems set the parameters for producing and accessing information. Interdependence among the components is governed by institutionalized norms and rules. The norms and rules are subject to international human rights agreements. For example, Article 19 of the *Universal Declaration of Human Rights* (UDHR) states that:

The inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world... Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.⁵⁶

Article 19 of the *International Covenant on Civil and Political Rights* (ICCPR) asserts that:

Everyone shall have the right to hold opinions without interference. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice. The exercise of the rights... carries with it special duties and responsibilities.⁵⁷

It is important to note that Article 29 of the UDHR also asserts that:

In the exercise of his [sic] rights and freedoms, everyone shall be subject only to such limitations as are determined by law solely for the purpose of securing due recognition and respect for the rights and freedoms of others and of meeting the just requirements of morality, public order and the general welfare in a democratic society.⁵⁸

International human rights law protects the rights and freedoms of the individual. It also insists on respect and responsibility for the rights and freedoms of others, making individual rights conditional. Again, value judgments as to what constitutes 'respect and responsibility' for others create marked differences in how rights are institutionalized and practiced. These international agreements bind states and are reiterated in regional and national human rights law – within democracies and within autocracies. Embedding human rights, duties and responsibilities in information ecosystems and the public sphere has been a continuous challenge historically, but new issues are being confronted as digital technologies are used to provide novel means of creating and circulating information and speeding up communication processes.

⁵⁴ Radsch (2023e, p. 1). This definition was developed in consultation with more than 40 practitioners and experts from around the world including many from the Global Majority World.

⁵⁵ Radsch (2023e, p. 1).

⁵⁶ UN (1948, Article 19).

⁵⁷ UN (1966, Article 19; emphasis added).

⁵⁸ UN (1948, Article 29; emphasis added).

4 Troubled Democracy and Mis- and Disinformation

As indicated, this report is concerned with what changes in information ecosystems and the public sphere mean for communication, for the integrity of information (a values-based judgment) and for the future of democracy. The *Global Digital Compact* defines information integrity as ‘access to relevant, reliable and accurate information and knowledge’, which is essential for an inclusive, open, safe and secure digital space where there is tolerance and respect in the digital space.⁵⁹ The Compact asserts that democracy cannot thrive if information ecosystems are prone to the ‘substitution of lies for factual truth’.⁶⁰

Our research assessment aims to understand how communicative processes work within information ecosystems and the public sphere. We understand *communication* to refer to the exchange of information between individuals or groups using shared concepts and signs, including direct conversations, commercial and public service media and as mediated by digital platforms. It is through communication that information is gathered and shared, voiced and heard. A well-functioning democracy needs effective communication, which depends on the availability of accurate information. As indicated, how information influences public opinion and decision-making hinges on fair and open communication within the public sphere. Some information might be helpful in contributing to knowledge that guides behavior in accordance

with social norms and that upholds fundamental human rights. As mis- and disinformation have gone viral there is a risk that helpful or useful information is crowded out or drowned out, increasing the fragility of democracies, jeopardizing human rights protections – people’s rights to freedom of expression, privacy, equality and justice, and compromising adherence to the rule of law.⁶¹

This is because democracies are based on a normative order that enables processes of societal self-determination. In democratic orders the public legitimizes complex norms and values, creating a fundamental structure for society defining how a state and its relations with other actors operates. The exercise of political authority and the distribution of goods and services depend on this ordering, which is coupled with narratives that legitimize and stabilize the normative order.⁶² These narratives are based on information and develop through communication processes. In well-functioning information ecosystems, those who are impacted by decisions taken within that order are assumed to play a role in defining the rules governing which decisions are taken. When the rules for automated systems and communicative practices are set by actors that are not perceived to be legitimate, decision-making processes become destabilized or corrupted.

It is in this context that the viral spread of mis- and disinformation as well as hate speech is depicted in the policy literature as ‘polluting’ the information ecosystem and threatening human progress (see Figure 1.2). However, this does not address wider questions about why this speech is so prevalent or who has the power to change the societal conditions that give rise to it, or the behavior of the big tech companies that facilitate its production and circulation.

⁵⁹ UN (2024b, paras 33, 34).

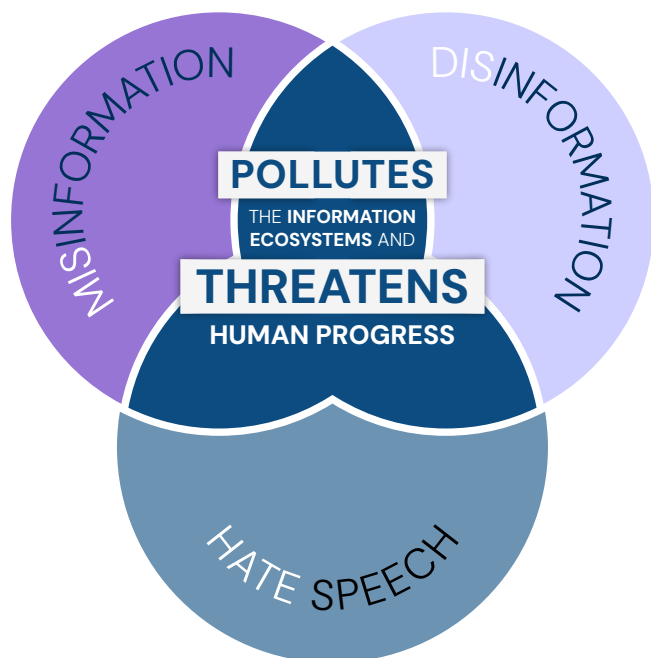
⁶⁰ Arendt (1968, p. 257).

⁶¹ Advox Team (2024), Global Voices supported by Deutsche Welle Academy (DW Akademie) and the Federal Republic of Germany through BMZ (Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung); see also Wagner *et al.* (2025).

⁶² Hamelink (2023); Kettemann (2022); Puppis *et al.* (2024).

Figure 1.2

Polluting the information ecosystem



Source: UN (2023a, p. 5)

Concern in policy and research communities typically centers around threats accompanying the declining ‘accuracy, consistency and reliability of information’.⁶³ But what counts as mis- and disinformation? A distinction between mis- and disinformation that is present in many policy documents is shown in Table 1.1, indicating that it is the intent to knowingly cause harm that is used to distinguish between them.

Table 1.1

Distinguishing between mis- and disinformation

	Awareness of falsity	Underlying intent
Disinformation	Aware	“Bad”
Misinformation	Unaware (“inadvertent”)	“Good / neutral”

Source: Bontcheva et al. (2020, p. 26)

In the Global North, research definitions of mis- and disinformation vary, although there is convergence around definitions in the policy literature. Definitions vary with respect to the nature and kinds of harm associated with mis- and disinformation and the level of specificity and granularity; on whether harm refers to individuals, groups, organizations or countries or to democratic processes; and on whether harm needs to be shown to have occurred.⁶⁴

Research on mis- and disinformation often does not mention hate speech. This may be because ‘hate speech’ is subject to the strong protection of free speech rights in some jurisdictions,⁶⁵ and there is no definition of it under international human rights law (although ‘incitement to discrimination, hostility or violence’ is prohibited under Article 20(2) of the ICCPR).⁶⁶ Mis- and disinformation may take the form of state-sponsored campaigns or government, anti-government or other political propaganda, or it may manifest through individual contributions. It can appear in legacy news media, online news media or the feeds of online services, and it can make use of numerous features of the infrastructure layer of information ecosystems.⁶⁷ In African regions, in India and in other countries in the Global Majority World,

⁶³ UN (2023a, p. 5). Misinformation refers here to unintentionally spread inaccurate information. Disinformation refers to knowingly false and intentionally disseminated information to cause serious social harm. Hate speech is that which ‘attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, color, descent, gender or other identity factor’ (UN, 2023a, p. 5 citing UN, 2019). Hate speech is included because of the way it pollutes information ecosystems and ‘threatens human progress’.

⁶⁴ Wardle & Derakhshan’s (2017, p. 15) influential report, *Information Disorder*, distinguished between disinformation – ‘information that is false and deliberately created to harm a person, social group, organization or country’; misinformation – ‘information that is false, but not created with the intention of causing harm’; and malinformation – ‘information that is based on reality, used to inflict harm on a person, organization or country’. The European Union defines disinformation as ‘verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm’ (EC, 2018, pp. 3–4). The concerns are about how ‘waves of false or misleading content can undermine social cohesion, cast doubt on factual information, and undermine trust in public institutions’ (OECD, 2024, p. 14). See also Altay et al. (2023a); EC (2020a); EC HLG (2018); François (2019); Kapantai et al. (2021); Möller et al. (2020); Ó Fathaigh et al. (2021); Wardle (2018).

⁶⁵ Gillespie (2020) for the United States, and see UN (2024e) and UN (2019, p. 2) United Nations Strategy and Plan of Action on Hate Speech which defines hate speech as ‘any kind of communication in speech, writing or behaviour, that attacks or use pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor’.

⁶⁶ UN (1966, Article 20(2)).

⁶⁷ For a list, see Bontcheva et al. (2020, pp. 45–46).

mis- and disinformation definitions are sometimes aligned with Global North definitions, but they are also distinctive, including varying practice regarding the sanctions that apply when illegal or harmful information is deemed to be present.⁶⁸

The *Global Digital Compact* calls on its member states to ‘commit to respect, protect and promote human rights in the digital space’ and to uphold international human rights law.⁶⁹ By 2030 the United Nations aims to:

Promote diverse and resilient information ecosystems, including by strengthening independent and public media and supporting journalists and media workers... Provide, promote and facilitate access to and dissemination of independent, fact-based, timely, targeted, clear, accessible, multilingual and science-based information to counter misinformation and disinformation... [and] Promote access to relevant, reliable and accurate information in crisis situations, to protect and empower those in vulnerable situations.

The Compact states that:

We must urgently counter and address all forms of violence, including sexual and gender-based violence, which occurs through or is amplified by the use of technology, all forms of hate speech and discrimination, misinformation and disinformation, cyberbullying and child sexual exploitation and abuse. We will establish and maintain robust risk mitigation and redress measures that also protect privacy and freedom of expression... [protecting] the rights of the child in the digital space, in line with international human rights law, including the Convention on the Rights of the Child.

It calls on the digital technology companies and developers to ‘co-develop industry accountability

frameworks... that increase transparency around their systems and processes, define responsibilities and commit to standards as well as auditable public reports’, including by providing researchers access to data:

To build an evidence base on how to address misinformation and disinformation and hate speech that can inform government and industry policies, standards and best practices... [including incorporating] safeguards into artificial intelligence model training processes, identification of artificial intelligence-generated material, authenticity certification for content and origins, labelling, watermarking and other techniques.

And it recognizes ‘the urgent need for strengthened data governance cooperation at all levels with the effective, equitable and meaningful participation of all countries and in consultation with relevant stakeholders to unlock the full potential of digital and emerging technologies’. It calls for ‘a balanced, inclusive and risk-based approach to the governance of artificial intelligence (AI)’.

There is clearly much to be done. As one participant in deliberations leading to the *Global Digital Compact* writes:

Although the final version of the *Global Digital Compact* saw a significant erosion of principles of equity, redress and commitments to international solidarity funding beyond AI, it is perhaps a triumph that sufficient consensus was reached with current geopolitical tensions and political polarisation to have anything to take forward at all.⁷⁰

There are scholars who argue that mis- and disinformation are not significant problems because the causal impact of these kinds of information on individuals is hard to demonstrate. It is also argued that mis- and disinformation is a small proportion of the information that people engage with. Others

⁶⁸ Africa Center for Strategic Studies (2024); Madan (2021).

⁶⁹ All quotes are from UN (2024b, paras 22, 32b, 35b, c, d, 31b, 32b, 36a, b, c, 38).

⁷⁰ Gillwald (2024).

argue that mis- and disinformation can ‘causally and adversely influence people’s beliefs, decisions, and behaviors’.⁷¹ The critical assessment of research in this report focuses on a range of approaches to the study of the threats and harms associated with information ecosystems, that is, those that consider the context and those concerned with identifying causal effects of mis- and disinformation on individual attitudes and behavior. We also attend to research on governance institutions and the practices on individuals and collective groups.

5 Limitations of the Report

This critical analysis of state-of-the-art of research on important components of information ecosystems is limited in several ways.

First, the analysis is structured principally around the three themes – news media, AI systems and data governance. It is limited in what it can reveal about the unequal material conditions of people’s lives and by the questions used to structure the analysis that focused on these three areas and on mis- and disinformation. Where research that is included highlights factors such as poverty and unequal socio-economic conditions leading to exclusions and discrimination, the report does discuss these issues as important contextual factors, but this is not the principal focus.

Second, this report focuses mainly on research on information and communication circulating within the upper service applications layer of information ecosystems, and not the network infrastructure layer. This layer is important in structuring information ecosystems. Although some attention is given to controversy about network neutrality policies and zero-rating data contracts, and to the

capacities of governments and internet service providers to shut down or block the internet, this report does not seek to address the substantial literature on internet governance.

Third, societies experience the ‘information crisis’ differently depending on their social, cultural, political and economic circumstances. The analysis in this report is limited by an imbalance in Global North and Global Majority World research sources that favors the Global North (as indicated in Section 1 of this chapter; see also Appendix: Methodology). This imbalance limits our analysis, and it is undoubtedly the case that we have tended to privilege Global North experience, and especially knowledge about the United States and Europe, notwithstanding our efforts to reach out to be more inclusive.⁷²

Fourth, this report was not designed to encompass the substantial field of research on ‘digital divides’. An effort was made to emphasize the distinctive experiences of information ecosystems in different parts of the world, and we acknowledge huge variations in the availability of meaningful internet connectivity and access as well the presence of restrictions on access to information.

Fifth, several other large bodies of research make only an occasional appearance in this analysis. This includes substantial research on cybersecurity, securitization, geopolitics and ‘digital sovereignty’ and the fields of the economic geography of digital labor markets or the (micro)economic analysis of digital markets.

Sixth, in attempting to cover broad fields of research, observations are made about ‘country’-level experience and institutional practices. The analysis was not designed to capture research on the micro level or specific sectoral experiences of information ecosystems. We do not include technology ‘use cases’ or detailed case studies of experience and practice.

⁷¹ Ecker *et al.* (2024b, p. 1), supported in part by the Australian Research Council (ARC), British Academy, UK Government, American Psychological Association (APA)/Centers for Disease Control and Prevention (CDC), Google and Google Jigsaw, European Research Council (ERC) and the European Commission Horizon project, Humboldt Foundation and Volkswagen Foundation.

⁷² A survey of research on digital platforms, for example, published in each of the years 2018 and 2021, confirms this bias in research in this area: countries in the lead were the United States, China and the United Kingdom; as a region the European Union had the largest proportion of papers (Ha *et al.* 2023), funded by the Ministry of Education, National Research Foundation and Ministry of Science and ICT (MSIT), South Korea.

Finally, the analysis in this report is inevitably limited by the fact that all research is guided by research questions selected for investigation by research communities, the funding available to do research, and the researchers' access to data.

6 Chapter Summary and Report Outline

It is important to acknowledge that 'scientific knowledge cannot be understood as absolute'.⁷³ This does not mean that uncertainty must lead to the conclusion that findings are arbitrary or unreliable. In the absence of certainty, a critical analysis of state-of-the-art research assessment can tell us what we can be reasonably confident about, what is controversial, and what the priorities should be for future research and policy action. The absence of certainty can create dilemmas for policy makers looking for ways to combat mis- and disinformation without abridging human rights commitments.⁷⁴ The chapters in this report yield an insight into why there is an information crisis, and what is or could be done to mitigate threats and harms associated with mis- and disinformation.

This report is structured to introduce readers to research on the integrity of information in the public sphere within the context of information ecosystems. Chapter 2 looks at how this is influenced by the news media industry's structure, its increasing dependence on digital platforms, and how declining trust in the news and practices that weaponize information are associated with political polarization. Chapter 3 tackles the way the integrity of information is influenced by AI systems developments and the implications for the protection of human rights and for democracy. Chapter 4 turns to how these developments – captured by the term 'datafication' – are seen from a political economy research perspective that focuses directly on the exercise of power by the big tech companies and the incentives they have in deciding how data is collected and used.

Chapters 5 to 8 then turn to what is or could be done to address the information crisis. Chapter 5 is concerned with public and policy makers' understanding of how information ecosystems are contributing to the crisis, and with an important response – literacy training. Here the focus is principally on research on measures to enable adults and children to protect themselves from harms associated with datafication and mis- and disinformation. Chapter 6 provides insights into the legislative and regulatory measures that are being taken to set rules and norms of behavior to change the strategies and practices of big tech companies when their business practices are misaligned with rights protections. Chapter 7 zeros in on a range of measures, from fact-checking to self-regulation to co-regulation, which are specifically intended to mitigate the harms of mis- and disinformation. In Chapter 8 the assessment turns to the steps being taken by a variety of individuals and groups to imagine and practice data governance in ways that are consistent with just outcomes for all. Chapter 9 concludes by summarizing key insights for researchers and lessons for both big tech companies and governments.

Here is a guide to the questions and research areas addressed in each chapter.

Chapter 2: News Media, Information Integrity and the Public Sphere.

This chapter examines what research tells us about changes in legacy and online news media, and what can be done to promote information integrity and a democratic public sphere. *What are the market structures in the news media industry, and the power relations between news media organizations and digital platforms? What is the relationship between news media, a healthy public sphere and democracy? What strategies are available to the journalism profession to work towards building trust in the news media?* The analysis includes research on the structural characteristics of news media markets and platformization,

⁷³ This does not mean there are no valid standards for making judgments about scientific evidence (Ecker *et al.*, 2024a, p. 30).

⁷⁴ Radsch (2022).

motivations to produce and consume mis- and disinformation and resilience, news media trust and distrust, the trustworthiness of legacy and online news outlets, news consumption and avoidance habits, the weaponization of information and political polarization.

Chapter 3: Artificial Intelligence, Information Ecosystems and Democracy.

This chapter examines research on the properties of AI systems (specifically machine learning algorithms) and their embeddedness in online content governance systems. *How is 'artificial intelligence' (AI) defined, and what are the relationships between AI systems development and internationally protected human rights? What are the interdependencies between AI systems development, the use of automated tools and democratic processes?* The analysis includes research on the relationships between AI systems and human rights, AI systems use and content governance (generation and moderation), and how these developments are related to changes in democracy, societal resilience and cohesion.

Chapter 4: Big Tech Power and Governing Uses of Data.

This chapter examines the relationships between the power of big tech companies and approaches to governing practices of data extraction and use – the processes of datafication. *What is the appropriate role of data and digital infrastructures within political communities? How are data aggregation and AI systems changing the way people build, share and receive information and knowledge? How do these big tech strategies and practices interfere with political deliberation, which is essential for the survival of participatory democracy?* The chapter provides an assessment of research in these areas and the

political economy of datafication processes. This includes research on digital infrastructure contestations, monopolization practices and business models, and the need to work towards democratic forms of data governance.

Chapter 5: Awareness of Mis- and Disinformation and the Literacy Challenge.

This chapter focuses on people's knowledge about the presence of mis- and disinformation in information ecosystems and literacy training initiatives enabling children and adults to identify these types of information and to protect themselves from their harmful consequences. *How aware are the public and policy makers of the risks and harms of mis- and disinformation? What are the approaches to media and information literacy, and AI literacy, and what is the evidence on their effectiveness?* This chapter provides an assessment of research in the context of the need to protect the fundamental human rights of both adults and children.

Chapter 6: Governing Information Ecosystems: Legislation and Regulation.

This chapter provides an account of selected legislative and regulatory tools that are available to governments to mitigate the harms of mis- and disinformation, and to govern the way mainly big tech companies operate. *What types of governance approaches are available? What approaches to information ecosystem governance are being promoted at the global level? What are some of the legislative, regulatory and judicial approaches to governing information ecosystems?* This chapter emphasizes normative goals and rules embodied in governance approaches, providing an insight into tensions between these goals and rules and their implementation, as reflected by the experience and interests of different actors. The analysis

focuses on principles and guidelines reflected in legislation and regulations with respect to network infrastructure, privacy and data protection, digital platforms, AI systems and news media.

Chapter 7: Combating Mis- and Disinformation in Practice.

This chapter looks in detail at specific governance measures to combat mis- and disinformation by civil society organizations and governments. *What content governance efforts are being made to combat mis- and disinformation? What are the challenges in achieving effective governance of information ecosystems? In what ways are human rights protections jeopardized by governance aimed at curtailing online mis- and disinformation? What is known about the public's appetite for interventions to moderate online mis- and disinformation?* The analysis emphasizes the need to differentiate between the stated aims of governance and its consequences when practice falls short of normative expectations. It focuses on fact-checking, industry self-regulation, co-regulatory approaches and what views are expressed by the public about how mis- and disinformation issues should be addressed.

Chapter 8: Towards Data Justice in Information Ecosystems.

This chapter examines how the monopolistic power of big tech companies creates biases and harmful discrimination and exclusions, and infringes on people's human rights in a data economy that thrives on data extraction and monetization. *Why do corporate incentives, strategies and practices involved in designing, developing, selling and controlling data lead to epistemic injustice? What strategies and tactics are individuals and communities developing to resist the extractive features of the data*

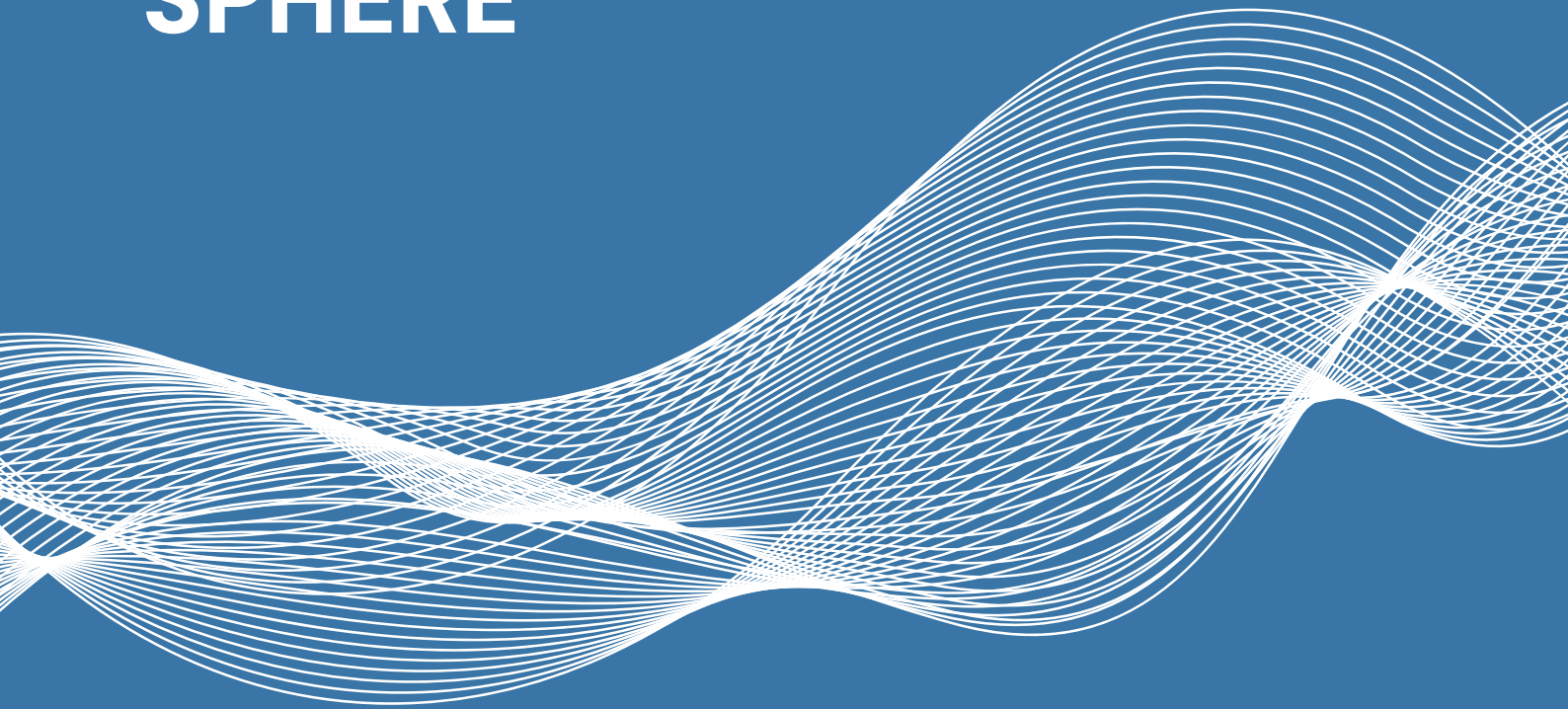
economy? This chapter emphasizes the individual and collective dependencies and inequities resulting from datafication, and how datafication practices can be reimagined to empower individuals and communities and contribute to data justice. It focuses on the consequences of biased AI systems for human rights guarantees and democratic decision-making, and individual and group (local, municipal and national) resistance strategies.

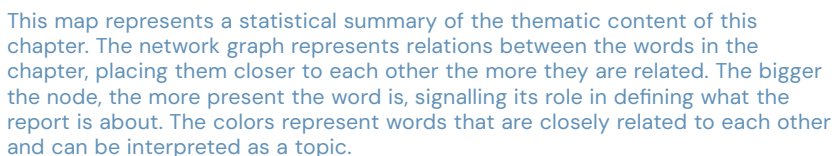
Chapter 9: Conclusion – Information Ecosystems and Troubled Democracy.

This chapter provides a discussion of the principal thematic insights that emerged from our assessment of state-of-the-art research, comments on key characteristics of the research we reviewed, a summary of each of the preceding chapters with key insights, and a brief account of the limitations of the report as well as a final word on what next. (See also the Executive Summary.)

CHAPTER 2

NEWS MEDIA, INFORMATION INTEGRITY AND THE PUBLIC SPHERE





The map is generated by the OID on the basis of the chapter's text using GarganText – developed by the CNRS Institute of Complex Systems. Starting from a co-occurrence matrix generated from chapter's text, GarganText forms a network where words are connected if they are likely to occur together. Clustering is conducted based on the Louvain community detection method, and the visualization is generated using the Force Atlas 2 algorithm.

[Link to the interactive map here](#)

This chapter examines what research tells us about the multiple causes and consequences of changes in legacy and online news media, and what can be done to promote information integrity and a democratic public sphere.¹ The chapter begins with a brief discussion of what is included as legacy and online news media.

The research synthesis focuses on:

- **What are the market structures in the news media industry and the power relations between news media organizations and digital platforms?** The discussion highlights research on the platformization of news, the dependence of news media on platforms and declining advertising revenues, and efforts to monetize news content that create incentives for the production and circulation of mis- and disinformation.
- **What is the relationship between news media, a healthy public sphere and democracy?** This briefly discusses how the normative expectations for the news media are conceptualized, interpreted and practiced in different parts of the world, emphasizing the need to take account of the Eurocentricity of much research in this area.
- **Who engages with news, and what factors account for whether people trust the news and how they perceive the trustworthiness of news media organizations?** This addresses changes in journalism standards and practices, what is known about the way audiences engage with news, their various uses of news and their reasons for avoiding news, their resilience to mis- and disinformation, and evidence on the way actors seek to weaponize information. Evidence on whether engagement with the viral circulation of mis- and disinformation should be treated as a principal cause of polarization of public opinion is also examined.
- **What strategies are available to the journalism profession to work towards building trust in the news media?** Research on measures to increase audience engagement with trustworthy news is briefly discussed.

The chapter provides an insight into a wide range of research traditions, looking at both the effects of news media engagement on people's attitudes and behaviors and the broader complex factors that influence how diverse information ecosystems are experienced.

Further discussion of the news media, politics and trust theme is found in Chapter 3 which examines how the news media industry is engaging with AI systems as part of this chapter's focus on AI systems, information ecosystems and democracy. The governance of legacy and online news media is examined in Chapters 6 and 7, and the role of non-mainstream news media in information ecosystems is examined in Chapter 8.

¹ For background reading, see Benequista et al. (2019); Couldry & Curran (2003); Curran & Hesmondhalgh (2019); Curran & Park (1999); Mano & Milton (2021); Wasserman (2018). See Appendix: Methodology for details of literature review process.

1 Introduction

This chapter starts with definitions. It is difficult to draw neat boundaries around what ‘news’ is and who can claim to be a ‘news producer’. Our concern here is primarily with news media organizations, although we also discuss the activities of individuals who produce mis- and disinformation and who are not affiliated with recognized news organizations.

Legacy news media – television, radio and offline newspapers – and online news media coexist today, and the many participants in the news media industry share norms such as the protection of sources and the goal of objectivity or impartiality. Digital journalism has come to refer to the ‘practices of newsgathering, reporting, textual production and ancillary communication that reflect, respond to, and shape the social, cultural and economic logics of the constantly changing digital media environment’.² News organizations may be commercial businesses relying on advertising revenues, state-owned, public service media (PSM), or collectively owned.

Legacy and online news media share characteristics including recording information with digital technologies; news formats that are intended to engage audiences with content; the production of content that can be accessed at any point in time or location; and an environment in which PSM and collectively owned smaller organizations struggle for prominence on digital platforms. Interaction between news organizations and those who engage with news content shapes news media agendas, although not with identical powers of decision-making. This relationship influences the perceived legitimacy of news media content, and

whether news outlets are perceived as trustworthy. This in turn depends on ‘affordances’ – that is, the instrumental and social features that result from users’ interactions with technology,³ and the governance arrangements that are applied differently to legacy and newer news producers.⁴

Some news media are designated as alternative media. This form of media goes by numerous labels: ‘radical’, ‘citizens’ media’, ‘advocacy journalism’, ‘participatory’, ‘independent’, ‘activist’ and ‘grassroots autonomous media’. These organizations differ substantially in how they position themselves in relation to mainstream news and in their political orientations.⁵ Alternative media are sometimes defined as ‘a range of media forms and practices, from radical critical media to independent media, and from grassroots autonomous media to community, citizen and participatory media’.⁶ Some position themselves as ‘counter-hegemonic’, emphasizing non-commercial amateur production and limited financial resources, while others focus on the use of technology for emancipatory purposes.⁷

This is the context in which much research focuses on whether the news media are trustworthy and whether news media audiences trust the content they encounter. ‘Trust’ is not always defined or operationalized in the same way, but it is generally associated with an:

Individual’s willingness to be vulnerable to media objects, based on the expectation that they will perform a) satisfactorily for the individual and/or b) according to the dominant norms and values in society (i.e., democratic media functions).⁸

² Burgess & Hurcombe (2019, p. 360). The European Union’s Media Freedom Act of 2024 defines a ‘media service’ as one where the principal purpose is ‘providing programmes or press publications to the general public, by any means, in order to inform, entertain or educate, under the editorial responsibility of a media service provider’ (EC, 2024b, Article 2(1)). It is unclear whether this excludes individual journalists, bloggers, non-profit news website and other organizations. The Council of Europe and other human rights organizations employ a broader definition. Equally hard to define is what privileges and protections the news media should enjoy despite being protected by international law, as changing technologies mean that claims to these privileges are disrupted, and it is difficult to hold a broadly defined ‘news industry’ to traditional normative professional standards; see Seipp *et al.* (2023a); Tambini (2021).

³ For definitions of ‘affordance’ as used in the research literature, see Evans *et al.* (2017); Hopkins (2020); Ronzhyn *et al.* (2023); Neubaum & Weeks (2023), supported in part by the Ministry of Culture and Science of the German State of North Rhine-Westphalia and by Stiftung Mercator.

⁴ EC: Directorate-General for Communications Networks *et al.* (2022); see also Mazzoli (2020), supported by the Economic and Social Research Council (ESRC), UK.

⁵ Cushion (2021), supported by Ofcom, the UK communications regulator.

⁶ Jeppesen (2016, p. 54).

⁷ Harlow (2017).

⁸ Fawzi *et al.* (2021, p. 156); see also Strömbäck *et al.* (2020), funded by Riksbankens Jubileumsfond for the Advancement of the Humanities and Social Sciences, Sweden.

Some research on news media and trust focuses on individuals' trust. Other traditions focus on relationships between actors, recognizing that 'trusting is not a matter of blind deference, but of placing – or refusing – trust with good judgement ... [and] we need social and political institutions that allow us to judge where to place our trust'.⁹ When the news media are seen by publics as being untrustworthy, this contributes to the undermining of democracy.

The history of asymmetric global and regional news media markets and news flows between the Global North and Global Majority World or 'non-aligned' countries was studied long before the internet and debates about mis- and disinformation. The term 'propaganda' was used in analyses of the hegemony of news organizations mainly in the Global North.¹⁰ Digital platforms and the platformization of news media are now seen as weakening news media organizations and contributing to declining trust in information ecosystems around the world, with accompanying threats to democracy, as violent and toxic discourses are amplified online.¹¹ Combined with news organizations that in many countries are heavily dependent on advertising revenues and face declining revenue due to competition from the digital platforms in the ad tech market, the news industry is in crisis in many parts of the world.¹²

In addition, for journalists and news media organizations in the Global Majority World (and especially lower-income countries), a digital connectivity gap influences whether journalists and their audiences are able to benefit from online journalism practices, and high-choice news media environments are not available everywhere. Access to online news can be limited by weak or absent internet availability and affordability, the high cost of digital services and poor access to journalism training. Local reporting norms vary by country, and

journalists' safety is often at risk.¹³ Various practices limit or discourage online access to certain kinds of information, for example 'zero-rating' data-pricing policies of network operators and the use of algorithmic personalization tools that lock people into 'walled garden' environments, where it is more costly for them to access diverse sources of news that are not part of the package permitted by their online data service contracts.¹⁴ These conditions influence what information people are exposed to online.

The next section examines research on the structure of the news media industry and the power relationships between news media organizations and the big tech companies' digital platforms.

2 News Media and Structural Power

The structure of the news media industry influences how news in different locations around the world is organized. Market structures of the privately owned news industry involve power relationships that create different economic and political incentives for the production and circulation of news content. These influence whether news content producers and news content are trusted, and they are visible in ownership conditions, the extent to which news organizations are politically independent, and in the relationships between digital platform companies that increasingly host news content. Research on news media trust often focuses on whether media power and influence are concentrated in ways that limit the diversity and plurality of news media content. The structure of the privately owned news media market impacts on the viability of PSM, which

⁹ O'Neill (2002, p. 7); see also Blöbaum (2016); Frislich & Humprecht (2021).

¹⁰ The non-aligned movement is a group of 120 mostly low- and middle-income countries active from the 1950s that elected not to align themselves with or against any major power bloc and that remains active; see Mansell & Nordenstreng (2006); Vincent & Nordenstreng (2015).

¹¹ Benkler (2020); Lasswell (1971); Rantanen (2024); Thussu (2022).

¹² Recuero (2024), supported in part by the National Council for Scientific and Technological Development (CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil.

¹³ Conroy-Krutz & Koné (2022), independent pan-African research network, supported by National Endowment for Democracy (NED), an NGO, US; see also Chiumbu & Munoriyarwa (2023). Fixed broadband networks are relatively rare in many countries in the Global Majority World, where people tend to rely on the mobile internet for access to online news.

¹⁴ Palmer & Toff (2022), supported in part by Google UK as part of the Digital News Initiative; see also Aharoni *et al.* (2021). Service contract fees can inhibit citizens' willingness to consume news from a broad range of sources (zero-rating and network neutrality issues are discussed in Section 4.1, Chapter 6).

also experience the platformization of their news content with impacts that vary with their funding arrangements.

Studies on structural power in the news media industry are typically conducted at country level, although differences are also examined between types of news producers that operate sub-nationally.¹⁵ Research in the political economy tradition focuses on how media power is exercised within the news media industry and the way news media create 'the terrain for other actors to contest power'.¹⁶ Thus:

Concentrated media power ... is antidemocratic both because it hands definitional, analytical, and interpretive power to unelected organizations and because it undermines the ability of citizens to acquire and exchange the information and ideas necessary to make informed decisions about public life. It is also dangerous, because it distorts the logic of the media industries themselves, transforming them from vehicles of symbolic interaction to increasingly significant engines of capital accumulation.¹⁷

Research on news media in the political economy tradition focuses on asymmetries of power between those producing and/or circulating news content and individuals or groups. Structural asymmetries are assumed to be present because of the power of dominant news media producers – and big tech companies – to control how audiences are exposed to news, that is, how corporate priorities for profit from advertising and the monetization of data generated by the audience's online interactions support the deployment of algorithm-driven news personalization systems.¹⁸

The dominance of big tech companies creates pressures on the news industry to change its operations and organizational frameworks: 'the rise of digital technologies, in a neoliberal, political, and economic climate, has facilitated the "platformization" of infrastructures and the "infrastructuralization" of platforms' (i.e., the ubiquity of digital platforms in people's lives).¹⁹ In this context, digital platforms are akin to publishers, and even editors, as news production relies on the algorithms, advertising markets, data and content moderation standards of the platforms, although they resist being designated in this way.²⁰ 'Captured' by the digital platform companies, some news outlets become dependent on financial arrangements, while platform owners argue that they financial arrangements the news organizations through their public relations campaigns and informal relationships.²¹ There are signs that some of the big tech owners of platforms are becoming less interested in hosting online news as they turn to new sources of revenue growth from the integration of AI tools into their systems. This is likely to create additional problems for news publishers as they try to build interest in their own online sites and attract readers using subscription packages and limited advertising revenue.

Mis- or disinformation is driven by platform company profit motives and exploitation of the affordances of platforms by professional persuaders.²² The actors engaged in producing and circulating this information take advantage of the algorithm-driven ad tech market that engages in 'digital deceit' to amplify content.²³ This generates significant revenue for the platforms, for advertisers willing to have their content appear alongside this content, and for individual influencers (e.g., celebrities and creators of fake accounts and information).

¹⁵ For a special issue on how platform power is theorized, see Nieborg *et al.* (2024). See also Nielsen & Ganter (2022). For a comparative analysis, see Nielsen & Fletcher (2023), supported by Google UK as part of the Digital News Initiative; Freedman (2014, p. 324).

¹⁶ Freedman (2014, p. 324); see also Thussu (2022).

¹⁷ Freedman (2014, p. 327).

¹⁸ See Mansell & Steinmueller (2020); Wasko *et al.* (2011); Winseck (2022).

¹⁹ Plantin *et al.* (2018, p. 298); see also Garcoa Ramirez (2021).

²⁰ In the United States, Section 230 of the Communications Decency Act of 1996 exempts platforms from liability for content they host on their platforms in the interests of upholding free speech rights; there is ongoing controversy about whether this should change, and in this sense, the platforms resist designation as 'editors' (Cramer, 2020). In the European Union, under the Digital Services Act of 2022 – which refers to an earlier e-commerce directive – platforms (intermediary services) are not liable for the content they host if they act as a 'mere conduit' Chapter 2, Article 4; provide temporary storage 'caching' Chapter 2, Article 5; host without knowledge of illegal content and act quickly to remove or disable illegal content when they obtain such knowledge, Chapter 2, Article 6; and there are other provisions (EC, 2022c). Liability provisions vary in countries around the world.

²¹ Greene (2018); Nechushtai (2018); Nieborg & Poell (2018); Papaevangelou (2023); Radsch (2023b).

²² Bakir & McStay (2018), funded by the Arts and Humanities Research Council (AHRC), UK.

²³ Ghosh & Scott (2018); Pielemeier (2020).

Monetization of mis- and disinformation.

Websites that repeatedly published mis- or disinformation generated USD 2.6 billion in advertising revenue in 2021 worldwide (the United States accounted for more than half). Meta generated at least USD 30.3 million in ad revenue from networks it removed from its own platforms for engaging in coordinated inauthentic behavior.²⁴

The digital platforms' control over the advertising market gives them the power to dictate financial terms to news organizations.²⁵ Their leverage over the news media industry comes from their position as intermediaries between news media content producers and their publics, and their capacity to deprioritize news.²⁶ Platform dominance and news media organization dependency is due to big tech companies' ability to aggregate end users, which magnifies network effects.²⁷ Many news media organizations have shifted to monetizing content by tailoring their news to platform affordances (technical characteristics and rules of operation) to boost user engagement and advertising revenue.²⁸ Reliance on ad tech metrics (e.g., clicks/impressions as a news performance indicator) has created a more competitive newsroom culture, but also increased management surveillance.²⁹

Advertising on platforms has attracted traffic for some news publishers, but has not always translated into economic sustainability for their businesses.³⁰ Some news media organizations have transitioned to subscription models, owning online news distribution and hosting paid-for events, and to native advertising (ads with the look and feel

of content they appear with) hosted at their own news sites.³¹ Some benefit from direct payments by platform companies, for example Google or Meta, a compensation for hosting news content.³² There have been clashes over such payments and about how to value news. Some platforms such as Meta have threatened to remove news content, for example in Australia and Canada, and in some cases take action to do so when agreement is not reached.³³ Market concentration in the news media industry and in the platform market undermines democracy because of the way it distorts news organizations' capacity to contribute to a healthy public sphere.³⁴ Table 2.1 highlights tensions in the relationships between some of the largest digital platforms and news media organizations.

Table 2.1
Tensions between digital platforms
and news media organizations

Digital platform	Relationship between the platforms and news media organizations
X (formerly Twitter)	<ul style="list-style-type: none"> • Elon Musk's takeover of Twitter was followed by platform changes that were perceived as problematic/unfavorable for the journalism profession. • In January 2024, X suspended an unknown number of prominent accounts that were critical of the Israeli government, blaming the spam algorithms. • In October 2023, X stopped showing headlines in previews to improve the aesthetics of the iOS app. The measure of excluding titles from previews of links might have been meant to discourage users from sharing third-party content from media organizations. • Focusing on the French media sector in a six-month period after Musk's takeover, it was found that journalists started to question the broader legitimacy of social media as a journalistic tool, but engaged in 'strategic disconnection' instead of abandoning the platform.

²⁴ Elliott (2022); Skibinski (2021).

²⁵ Bell *et al.* (2017); Garcia Ramirez (2021); Nielsen & Ganter (2022); Radsch (2023).

²⁶ Kristensen & Hartley (2023), supported by the VELUX FONDEN, Denmark; Nielsen & Ganter (2022); Poell *et al.* (2023).

²⁷ Montero & Finger (2021); Nieborg & Poell (2018); Nielsen & Ganter (2022). For a comprehensive discussion of how platform dominance has been achieved, mainly from a Global North perspective, see Bannerman (2022); Moore & Tambini (2018, 2021), supported by the Social Sciences and Humanities Research Council (SSHRC) of Canada. For African and Latin American countries, see Mabweazara *et al.* (2020); Mabweazara & Pearson (2024).

²⁸ Bell *et al.* (2017).

²⁹ Petre (2021).

³⁰ Nieborg & Poell (2018).

³¹ Meese & Hurcombe (2021), supported by the Australian Research Council (ARC).

³² Nielsen & Ganter (2022).

³³ Meese & Hurcombe (2021).

³⁴ Bimber & Gil de Zúñiga (2020); Vaidhyathan (2022).

Digital platform	Relationship between the platforms and news media organizations
Facebook/ Meta	<ul style="list-style-type: none"> Research on the 'Facebook problem' includes critiques of technology and design; amplification of misinformation; abuses of market power; monopolization of the digital advertising market, undermining financial support for journalism; encouraging sensationalist journalism and clickbait; and limitations of self-regulation. Small tweaks to the News Feed's algorithm can have a profound impact on the visibility of news content.
Google	<ul style="list-style-type: none"> There has been criticism of the monopolization of online advertising markets. Google's news services tend to steer revenue toward the largest publishers. Google has developed products that enable off-site publishing in new formats, such as Google Accelerated Mobile Pages. This implies a loss of control over channels of communication and increased dependence on platforms as news intermediaries. Risks include losing control over editorial identity (search algorithms shape the way users interact with news content) and access to data (more detailed analytics are available on-site than off-site).
YouTube/ Google	<ul style="list-style-type: none"> There is concern about changes in YouTube's policies regarding the demonetization, delisting and removal of videos, with implications beyond the performance of individual videos.
TikTok	<ul style="list-style-type: none"> There are concerns about the virality of content directed at young people, the decline of legacy news as gatekeepers, and how its algorithm exploits children's vulnerability and distributes racist and sexist content (in addition to geopolitical security concerns between the United States and China).
Social media platforms in general	<ul style="list-style-type: none"> There is criticism of dominant social media platforms and their failure to moderate harmful content at key moments (e.g., Gamergate (misogynistic online harassment campaign); the Rohingya genocide in Myanmar from 2016; the 2016 US presidential election, the Christchurch New Zealand Mosque shooting 2019); or Alex Jones' promotion of conspiracy theories and the rise of online revenge porn. There are problems with automated moderation and overstated claims of success; difficulty accounting for context, subtlety, sarcasm and subcultural meaning; and insensitivity to the use of duplicate content in different contexts, such as terrorist propaganda reposted in a journalistic context.

Source: Collated from scientific papers and media accounts.³⁵

News media organization dependency on platforms is especially severe in low-income countries where press freedom is limited or non-existent.³⁶ Local news organizations in the Global North have been hard hit as they move their content online. Competition for audiences is reducing local news stories to little more than 'clickbait' in some countries.³⁷ News 'deserts' have been reported in Europe and the United States as news organizations close at local and sometimes regional levels. Some argue, however, that focusing on legacy news media distracts attention from the marginalization of certain audiences from the public sphere that has occurred historically.³⁸

News media organizations have problems in accessing audience data, which compounds the power asymmetry with the digital platforms.³⁹ News reporting benefits from the ability to monitor audience interest, and this requires verification data. Journalists report difficulties in accessing accurate data and in interpreting the partial data they do receive that is biased to favor the platforms.⁴⁰ Instant online news production also undermines news verification processes and the ability of journalists to fact-check mis- and disinformation.⁴¹ In countries in Latin America, where financing often comes from philanthropists, legacy news media are pressurized to focus their reporting on the interests of their funders.⁴² As a result, legacy news media are struggling to maintain their audiences and the credibility of their news.

There is varied evidence of diminishing trust in news when it is obtained *via* platforms.⁴³ For example, across 47 markets and six continents in 2024, survey respondents expressing concern about

³⁵ Claesson (2023); Germain (2024); Gillespie (2020); Nielsen & Ganter (2022); Notley *et al.* (2020); Peters (2023); Pickard (2020c); Poell *et al.* (2023); Van Natta *et al.* (2023), the last two supported by the Ministry of Science, Innovation and Universities (MICIU, Ministerio de Ciencia, Innovación y Universidades), Spain and the European Commission.

³⁶ BBC Media Action (2021); García Ramírez (2021); Nielsen & Ganter (2022); Schot (2020), Free Press Unlimited, the Netherlands, an independent foundation. For an overview of the economics of the media industry, see Rohn *et al.* (2024).

³⁷ Tomaz & Trappel (2022).

³⁸ Usher (2023); Verza *et al.* (2024).

³⁹ Meese & Hurcombe (2021); Nieborg & Poell (2018); Nielsen & Ganter (2022). There are issues around fraudulent reporting of data and the way AI-generated news is infiltrating new sites. The United States Federal Trade Commission (FTC) is focusing on how this affects data on market share, the potential for sales growth and the expansion by AI companies into new markets, potentially creating further pressure on news provider finances; see FTC (2024).

⁴⁰ Dommett (2023).

⁴¹ Baron (2002); Wahl-Jorgensen & Carlson (2021); Ross Arguedas *et al.* (2022b), supported by the Meta Journalism Project; Himma-Kadakas & Ojamets (2022), supported in part by the Anders Foundation; Dierickx *et al.* (2023c), supported by the European Commission; Carson & Gibbons (2023), supported by Facebook.

⁴² Labio-Bernal & Romero-Domínguez (2022).

⁴³ Ross Arguedas *et al.* (2022c) supported in part by the Facebook Journalism Project; van Dijck *et al.* (2018a).

online ‘fake’ news increased to 59% of the sample although concern varied hugely by country – South Africa, 81%, India, 58%.⁴⁴

Interviews with news workers in Brazil, India, the United Kingdom and the United States indicate how platforms exert pressure on journalism practice and hamper audiences’ ability to distinguish between credible and non-credible news sources.⁴⁵ This complicates the realization of journalistic values associated with news media trustworthiness,⁴⁶ and raises concerns about declining news media editorial control.⁴⁷ News media organizations face the multiple challenges of maintaining editorial authority independent of platforms and governments, maintaining high-quality news standards, and delivering in-depth and diverse content.⁴⁸

In summary, analysis of news media market concentration and structural dependence on big tech platforms demonstrates why many legacy as well as online news media organizations are facing crises that threaten their sustainability, and this has consequences for the health of information ecosystems as well as the digital public sphere.⁴⁹

3 News Media, the Public Sphere and Democracy

The relationship between the news media and the public sphere in a context of platformization and in the face of the circulation of mis- and disinformation is complicated, and it is also strongly influenced by country political conditions. For

example, journalists and other actors – women and other minority or disadvantaged individuals or groups – are facing threats, violence and murder. This is occurring alongside increases in the production and circulation of mis- and disinformation (including hate speech). Since 1993, 1,701 journalists have been killed, according to UNESCO data, with 50% of the deaths occurring outside conflict zones.⁵⁰ Research demonstrates that in relation to women and their rights: ‘misinformation, disinformation and defamation are real and pervasive threats ... and women tend to be targeted more frequently than men’.⁵¹ In the Global Majority World (and elsewhere), ‘the act of gaining voice and attempting to have influence in the public sphere, the act of asserting their own humanity and right to exist qua human beings, *takes place during ongoing deprivation, campaigns of disinformation, police brutality and/or military atrocity*’.⁵² For this reason, it is important not to lose sight of the material conditions of people’s lives when the focus is on the role of the news media and how to combat mis- and disinformation.

News media organizations are essential to meet the public’s need to be informed about matters of public value. Ideally they contribute to democracy by helping to foster community-building, enabling communication flows among members of society, keeping citizens up to date with events and by educating them.⁵³ News media organizations and professional journalists are expected to uphold normative goals. Article 19 of the *Universal Declaration of Human Rights* (UDHR) (1948) states that: ‘everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers’.⁵⁴

⁴⁴ Newman *et al.* (2024), core funded by the Thomson Reuters Foundation and a wide range of others, including academic, foundation, non-profit and industry partners.

⁴⁵ Ross Arguedas *et al.* (2022c), supported as above.

⁴⁶ Van Dijck *et al.* (2018a).

⁴⁷ Eichler (2023); Nielsen & Ganter (2022); van Dijck *et al.* (2018b); see also Hartley *et al.* (2023), supported by the VELUX FONDEN, Denmark.

⁴⁸ Eichler (2023, p. 283).

⁴⁹ See Nicholson (2024) on the political economy of media industries (this issue is addressed further in Section 4.5, Chapter 6).

⁵⁰ UNESCO (2024); at the time of writing, in August 2024.

⁵¹ Gallagher (2023, p. 58).

⁵² Banaji (2024, p. 13; emphasis added).

⁵³ Anderson (1983); Hanitzsch & Vos (2018).

⁵⁴ UN (1948, Article 19).

Article 19 of the *International Covenant on Civil and Political Rights* (ICCPR) (1966) similarly states that: ‘everyone shall have the right to hold opinions without interference ... [and] the right to freedom of expression’.⁵⁵ The ICCPR recognizes that these rights entail duties and responsibilities. These normative goals are ‘not self-executing’, and especially not in countries where democracy is fragile, or in authoritarian states.⁵⁶

The rights and responsibilities of the news media are contested especially when they conflict with the goals of actors who seek to secure power and privilege.⁵⁷ Studies of the news media’s role in democracies are often grounded in a Western understanding of how normative goals should be interpreted in practice. Beyond the West, scholars frequently insist that the rights embodied in international declarations and covenants must be interpreted through the prism of their own cultures. Human rights norms may be universal, but there are many ways they can be respected through the presence of news media that aspire to these norms, even if the news organizations do so in a variety of ways. This is especially so in Global Majority World countries where local practices differ from those in the Global North.⁵⁸ Failure to acknowledge this is symptomatic of Eurocentricity, which too often characterizes knowledge production, and this is present when ‘particular dominant social, political or economic interests’ influence how the news media industry operates.⁵⁹

International human rights declarations and covenants set normative goals for signatory countries. The ideal in liberal democracies is understood to mean that the news media’s role is to voice the concerns of the public and hold the powerful to account – ‘speaking truth to power’. Democratization is expected to be accompanied

by an independent media industry. A vibrant public sphere (or healthy information ecosystem) is central to the ideal of rational democratic deliberation. In this context, news media are expected to provide factual, accurate and impartial (or objective) information, although this view can be challenged when it is inconsistent with inclusivity.⁶⁰

In practice there are multiple co-existing public spheres, and people participate in public life with unequal power, often as counter-publics. This is especially so in the Global Majority World when people are seeking inclusion on the margins, which is a legacy of colonialism. This means that the news media cannot be expected to inform a singular public or operate as the only source of information when there are many sub-audiences to whom news producers can appeal.⁶¹ Thus, the concept of a democratic ‘public sphere’ is a normative ideal. Historically, and today, there are ongoing struggles to achieve the ideal of news media independence and impartiality, especially in the face of overtly illiberal conditions. This is not the least because in practice there are major issues around how critical of government the news media can be, in both the Global North and the Global Majority World.⁶²

Notwithstanding variations in practice, inclusive information flows are crucial because they influence the quality of public discourse and the formation and legitimacy of public opinion. If information is misleading or wrong, public discussion cannot be fair, and the quality of democratic discourse suffers: for example, ‘hate in the space in which we debate publicly is one of the main ways of weakening democratic institutions’.⁶³ It is therefore important to recognize that while the platform ‘algorithms segregate and personalize ... *they cannot on their own, explain entity divisions*’, that is, disputes

⁵⁵ UN (1966, Article 19).

⁵⁶ Milton & Mano (2022, p. 35); see also Hamelink (2023).

⁵⁷ Repucci & Slipowitz (2022) supported by Google Inc., the Hurford Foundation, Jyllands-Posten Foundations, Lilly Endowment Inc., Meta Platforms Inc., and National Endowment for Democracy; see also Mukhudwana (2021); Pintak & Ginges (2008) and Pintak & Nazir (2013), both part-funded funded by the Rockefeller Brothers Fund US; Romano (2013).

⁵⁸ Chasi & Rodny-Gumede (2022); Wasserman (2020a).

⁵⁹ Willems (2014b, p. 418).

⁶⁰ Bennett & Kneuer (2023); Dahlberg (2014); Devenney (2009); Habermas (2015); Hallin & Mancini (2012); Jungherr & Schroeder (2021); Rugh (2000); Schudson (1978); Wasserman (2020b); see also Schlesinger (2020), supported by the Arts and Humanities Research Council (AHRC), UK. The concept of the public sphere has been criticized for its assumption of ‘critical-rational publics’ (Gerbaudo, 2022). Habermas (2022) stresses that it is crucial to distinguish between the normative conditions for a democratic polity, where participants struggle to secure the rights to which they are entitled, and the empirical reality of exclusions and marginalizations.

⁶¹ Dutta & Pal (2020); Fraser (1992); Fraser & Nash (2014).

⁶² For a discussion on how human rights and democracy can be united in governance structures, see Besson (2011).

⁶³ Aruguete & Calvo (2023); Zuazo & Aruguete (2021), no page numbers due to our translations.

over accuracy and ‘truth’ within and among groups in society. Polarization in the public sphere ‘exists before and beyond’ the algorithm-driven personalization of news content that audiences find online.⁶⁴

In the wake of platformization and the proliferation of mis- and disinformation, news media organizations face a variety of challenges. These are experienced differently in countries around the world, even as journalists and news media organizations benefit from hosting news on digital platforms and can reach new audiences. These changes have implications for whether the news media are seen as trustworthy and whether news consumers trust the news they encounter, both on- and offline. Prior to the platformization of news media, history is replete with examples of partisan (and political party-funded) news media voicing the concerns of their segmented audiences, not the public as a whole. This has varied from country to country and with the extent to which PSM have been able to serve the needs of the public in an impartial way.

4 Trust in News Media

Research consistently finds that Western countries are experiencing a decline in trust in legacy news media – trust in journalism as an institution – but this is not declining in all countries or at the same rate; trust in news media has always varied among countries and news media organizations. The issue is the extent to which platformization and the structure of the contemporary news media industry is contributing to a decline in trust in news media content and in the trustworthiness of news media organizations.

In the Global North, people’s news consumption habits have been steadily moving from legacy

news media to online sources and social media platforms.⁶⁵ A study in 2022 compared people’s trust in news on a range of digital platforms, including Facebook, YouTube and Google in Brazil, India, the United Kingdom and the United States. It found that trust depended on the platform, the country, the audiences and the kinds of news, but also that these sources were less trusted than legacy news media.⁶⁶

Brazil has seen one of the steepest declines in news media trust, dropping from 62% to 43% between 2015 and 2024, with the far right playing a key role in growing distrust of legacy media. In the West, some countries do not seem to be affected by declining levels of trust. In Denmark, trust is relatively stable, at around 57%. In South Korea, trust is reported to have risen from 22% to 31% between 2016 and 2024. There are countries where trust is high and increasing. In Kenya, it went up from 50% to 64% between 2020 and 2024, and in Thailand, from 50% to 54% between 2021 and 2024. In the United States there is a strong and asymmetrical decline in trust in news media between left-wing and right-wing voters, with a similar pattern in some countries beyond the West.⁶⁷

It is important to keep in mind that not everyone accesses the news, and that media trust can be associated with a media element, such as a person (a journalist, an expert), a source (e.g., *The Financial Times*, Fox News) or a type (television, radio, press), or it may be understood generally to apply to ‘the media’.⁶⁸ In addition, the results of surveys on media trust are questioned by some scholars who argue that it is unclear whether survey respondents understand what journalism standards are or should be when answering the survey questions.⁶⁹

Research on media trust tends to focus on overall trust in the news media, in news media as a public institution, in media organizations and their

⁶⁴ Political polarization is discussed in Section 4.4 in this chapter.

⁶⁵ See Ofcom (2023c).

⁶⁶ Mont’Alverne *et al.* (2022), funded by the Meta Journalism Project.

⁶⁷ Data for the named countries: Newman *et al.* (2024), supported by the Google News Initiative as well as multiple public and private funders. See also Newman *et al.* (2022), supported by a range of public and private funders, including BBC News, Ofcom and the Google News Initiative; Strömbäck *et al.* (2020), funded by Riksbankens Jubileumsfond for the Advancement of the Humanities and Social Sciences, Sweden; and Hanitzsch *et al.* (2018).

⁶⁸ Skovsgaard & Andersen (2020).

⁶⁹ Bernardi & de Moraes (2021); Bhat & Chadha (2020); Christoforetti & Becker (2023).

ownership, in media types (television vs. radio), in specific outlets, in media coverage and/or trust in journalists.⁷⁰ Trust in ‘media’ is used interchangeably with trust in the ‘news media’, in legacy media and sometimes in online media, which leads to ambiguity.⁷¹

Quantitative survey methods are used to reveal levels of trust among individuals, while other research focuses on industry structural factors that influence trust in news. Both face the problem of ambiguities around definitions of ‘trust’ and ‘news media’.⁷² Research reveals associations between trust and individual factors (socio-demographic, political and social attitudes) based on aggregate data (e.g., at country level) or individual-level data. Most studies measure overall trust in news media, which tends to measure trust in journalism as an institution, and are based on self-reports that may not be indicative of how people behave. Trust is also studied using qualitative methods that provide a deeper insight into why people trust or distrust news.

The results of research on media trust do confirm concerns about how the news media are implicated in increasing polarization within and between individuals and groups worldwide, but they do not provide clear answers as to exactly how they are implicated. To explain why this is so, we need to understand the concepts and theories about the impact of the media on individuals and societies that are present in research that informs studies of trust in the media – declining or otherwise.

4.1 CHANGING JOURNALISM STANDARDS AND NEWS MEDIA PRACTICES

Substantial resources were needed historically to produce and disseminate news, and legacy news media functioned as ‘gatekeepers’, selecting

what they deemed to be important.⁷³ When this power was accompanied by adherence to widely promoted standards of reporting (e.g., accuracy, impartiality), this was seen as a positive contribution to democracy, to information integrity and healthy information ecosystems. When the news media excessively amplifies certain narratives, this can contribute to democratic fragility. In some countries (e.g., Brazil, France, Italy, Spain and the United States), for example, the news media amplify far-right narratives when they report the discourse of far-right populist figures on issues such as immigration, foreign affairs, the environment or gender discrimination.⁷⁴ Research on the media landscape during the 2016 United States presidential election, for instance, suggests that the propagation of mis- and disinformation took advantage of structural weaknesses in the country’s media institutions.⁷⁵

Digital platforms are said to operate as a new ‘fifth estate’ because they have a gatekeeping role, although they resist designation as ‘media’.⁷⁶ This may lead to the presence of more diverse voices in the public sphere, but it raises questions about journalistic values.⁷⁷ During the two world wars, some news organizations sought to distance themselves from state propaganda, claiming to emphasize ‘facts’ and ‘objectivity’, especially in the United States.⁷⁸ However, in other circumstances, such as the ending of Apartheid in South Africa, the news media have faced difficult trade-offs – between encouraging the new democratic government and criticizing its actions.⁷⁹ What accuracy and ‘truth’ mean is understood differently in countries around the world, and procedures for achieving accurate reporting are being challenged as digital technologies contribute to changes in journalism practice.⁸⁰ Journalism is positioned in the literature as co-evolving with social media platforms

⁷⁰ Fawzi *et al.* (2021).

⁷¹ Schranz *et al.* (2018).

⁷² Christofolletti & Becker (2023).

⁷³ Shoemaker & Vos (2009).

⁷⁴ Pérez-Curiel *et al.* (2021).

⁷⁵ Benkler *et al.* (2018).

⁷⁶ Dutton (2023).

⁷⁷ Tandoc Jr & Vos (2016).

⁷⁸ Schudson (2022); Tuchman (1972).

⁷⁹ Wasserman (2020b).

⁸⁰ Habermas (2022).

such that practice ‘not only symptomatically reflects, but also pragmatically adapts to and influences the changing media environment’.⁸¹ Thus, digital journalism should not be understood as ‘journalism that is transformed by being digital’, but instead as a practice that increasingly embodies the use of digital technologies.⁸²

A meta-analysis of research published between 2013 and 2018 indicates a shift in focus from studies of how digital technologies impact on journalism to how journalism reflects and impacts on society.⁸³ The discourses, practices and logics of journalism shape the cultures, technologies and products of news media. As an agent of change, ‘digital journalism’ is seen as influencing the status and role of digital platforms (e.g., the legitimization of X/Twitter as a ‘global newsroom’) and as triggering changes in technical processes and practices (e.g., the growing popularity of news content on Facebook led Meta to acknowledge its editorial and curatorial role, and responsibility for the content it fosters, which then led to changes in its algorithms and to greater efforts to signal contested news, harmful content and mis- and disinformation on its platforms).⁸⁴

In the digitized news environment news organizations risk losing editorial control. This can diminish their credibility and lead to perceptions of news media bias. A competitive journalism culture within newsrooms, fueled by scrutiny of performance metrics and managerial surveillance, intensifies pressures on journalists and is widely seen as leading to a deprioritization of investigative journalism. There is evidence, however, of positive outcomes when journalists take advantage of digital services; for example, the use of WhatsApp in Rwanda has helped journalists to extend their coverage and educate each other through debate about their practices.⁸⁵

In addition, ‘alternative media’ organizations and journalists may be co-opted by the far left and positioned as criticizing commercial values, while right-wing media is likely to be associated with alleged ideological partisanship.⁸⁶ Both function as counterpoints to a dearth of diverse viewpoints, yet alternative news media is often said to engage in ‘one-sided and ideologically motivated “campaign” journalism’.⁸⁷ These news media do play a role in fostering dialogue, enabling marginalized voices to be heard and challenging the status quo, even if some are involved in circulating exclusionary narratives that may contribute to audience polarization.⁸⁸ Unfortunately, most studies of news diversity exclude alternative media, even where it operates as an influential competitor to legacy media. In research on media trust, and regardless of which type of news media is studied, assumptions must be made about how engagement with news content influences attitudes and behaviors.

4.2 NEWS MEDIA TRUST AND AUDIENCE ENGAGEMENT

Some studies of the impact of mis- and disinformation circulated by the news media seek to identify how news or information exposure can directly cause changes in attitudes and behavior by isolating news media impacts from other factors.⁸⁹ For example, the ‘hypodermic needle model’ (sometimes known as the ‘inoculation model’) of media effects suggests that information will trigger a similar reaction in everyone exposed to it, regardless of people’s characteristics.⁹⁰ It was initially developed to understand the effects of government propaganda in the era of mass media. This approach grants little or no agency to people and their ability to interpret the information.

A ‘two-step flow model of communication’ was later developed to add context, positioning opinion

⁸¹ Burgess & Hurcombe (2019, p. 360).

⁸² Duffy & Ang (2019, p. 378); see also Zelizer (2019).

⁸³ Steensen *et al.* (2019).

⁸⁴ Burgess & Hurcombe (2019, p. 360); for a systematic review of research on ‘data journalism’, see also d’Haenens *et al.* (2022); Erkmen (2023).

⁸⁵ McIntyre & Sobel (2019).

⁸⁶ Ihlebæk *et al.* (2022), supported by the Research Council of Norway.

⁸⁷ Ihlebæk *et al.* (2022, p. 1269), supported by the Research Council of Norway.

⁸⁸ Benkler *et al.* (2018); Siaper (2023).

⁸⁹ Anderson (2021); Klapper (1960); Lasswell (1971); McQuail (2010).

⁹⁰ Bineham (1988).

leaders as playing a role in mediating between news media and their audiences.⁹¹ A ‘selective exposure model’ proposed that people choose which news media to engage with based on their pre-existing views, assuming more limited news media effects.⁹² Other research focused on how the media shapes attitudes, suggesting that the media has a ‘cultivation’ role, that is, audiences tend to view the world as it is depicted in the media.⁹³ ‘Agenda-setting theory’ and ‘framing theory’ inform studies of the capacity of the media to set an agenda and to influence people’s selection of topics that matter to them.⁹⁴ These models of media effects are influential, and they benefit from new methods for measuring the effects of the information that circulates on social media platforms.⁹⁵

Other approaches to media effects are informed by theories from behavioral economics. Here the focus is on cognition and on the effects of nudging people away from mis- and disinformation based on understanding affective and cognitive responses. This work uses insights into cognition to provide cues to encourage people to change their online behavior, and is largely based on experimental studies. Nudging may aim to get people to attend to the accuracy of information. This assumes a ‘limited-attention utility model’ derived from the economic and psychological analysis of how choices are influenced by people’s pre-existing preferences, recognizing that cognitive capacities are limited.⁹⁶ Some of this research finds that average exposure to mis- and disinformation is not as high as is sometimes claimed, and that social media is not the primary cause of broader social problems, such as polarization. Exposure to false and inflammatory content has been found

to be concentrated within fringe groups with high motivation to seek this information out.⁹⁷

As early as 1996 it was concluded that ‘despite the volume of research, the debate about media effects – whether it can be shown empirically that the specific mass media messages, typically those transmitted by television, have specific, often detrimental effects, on the audiences who are exposed to them – *remains unresolved*’.⁹⁸ The search for the effects of mis- and disinformation continues in this tradition to discover ways to mitigate harms.

Other research traditions start from a different set of premises and have a similarly long history. The ‘audience research’ tradition, for example, is interested in how audiences interpret media content. This approach examines how people’s lives are ‘mediated’ by their relationships or engagements with information such as the news media.⁹⁹ It assumes that audiences have sufficient agency to interpret the news, and will do so in ways that are conditioned by their contexts. In contrast to media effects studies, both quantitative and qualitative methods are used, as in the case of ‘audience reception’ studies that seek to understand how audiences and the media co-produce information and cultures.¹⁰⁰ Studies may focus on the ‘uses and gratifications’ that audiences experience when they engage in news selection,¹⁰¹ and it is acknowledged that engagement (or non-engagement) with legacy and online media is important for people’s – and especially young people’s – ability to make sense of the world around them. Indeed, those who do engage online are depicted as living ‘inside’ media, and research may focus on how teenagers construct identities

⁹¹ Katz (1957).

⁹² Stroud (2017).

⁹³ Gerbner *et al.* (1980), supported by the Administration on Aging, Department of Health, Education and Welfare, US.

⁹⁴ Goffman (1974); McCombs & Shaw (1972); Valenzuela *et al.* (2023).

⁹⁵ Choi *et al.* (2020); Scott *et al.* (2022), supported by the Arts and Humanities Research Council (AHRC), UK.

⁹⁶ Pennycook & Rand (2022).

⁹⁷ For a review of the literature, see OECD (2022b); Pennycook & Rand (2022). Budak *et al.* (2024) calls for more research on exposure to content among extremists and fringe groups, and efforts to limit demand for this kind of information by curtailing political elites and legacy media that spread this information. Some authors in Budak *et al.* (2024) worked for Microsoft Research, some were participants in the US 2020 Facebook and Instagram Election Study, and the research was partly funded by Meta and Google Research.

⁹⁸ Livingstone (1996, p. 306, emphasis added).

⁹⁹ Mediation or ‘mediatization’ research is a longstanding research tradition on how people engage with and are influenced by offline and online information (Couldry & Hepp, 2016; Silverstone, 2007).

¹⁰⁰ Ong & Das (2020) point out that research on media effects is caught in a pendulum swing back to older assumptions of ‘hypodermic needle media effects’, which, they argue, is misleading in an era of datafication.

¹⁰¹ Livingstone (1998).

through online interaction, or why audiences chose their preferred media diets.¹⁰² The ‘audience research’ tradition emphasizes that audiences use technologies in unexpected ways, and that they engage actively with news media content. Here the focus is on what can be learned from studies of how individual attitudes and behavioral characteristics influence trust in information and news media.

Studies often examine individuals’ responses to news media, finding a variety of associations between individual characteristics and reported trust in news media. These studies rely on quantitative data collected at the individual level, finding, for example, that in Germany people with higher levels of interpersonal trust (the propensity to think that others will not harm them) report higher levels of trust in the news media.¹⁰⁴ In the United States, research finds that those with higher levels of political cynicism have less trust in the media.¹⁰⁵

The socio-demographic and age factors that are associated with news media trust are inconsistent across countries. In some countries men are less trusting, while the opposite is found in Israel, for example, and no association between trust and gender was found in the United States.¹⁰⁶ In Brazil, India, the United Kingdom and the United States, less educated people are found less likely to trust news media.¹⁰⁷ In contrast, a study across 44 countries found a slight decrease in trust with each additional year of schooling, and longitudinal research shows that those with a higher education degree are slightly less likely to trust the news media.¹⁰⁸ There is also evidence that diaspora communities tend to make greater use of non-mainstream media, while long-term residents make much less use of it.¹⁰⁹

Ideology and partisanship can influence trust in news media. This research is dominated by evidence from the United States, finding, for example, that Republicans are associated with significantly lower levels of trust in the media,¹¹⁰ and that trust in local and national news organizations has declined more rapidly for Republicans than for Democrats.¹¹¹ In other countries, trust is found to be more closely associated with attitudes towards extremism and populism than with left-right commitments. Some studies find that those who situate themselves in a more extreme ideological position are less likely to trust the media; others that those with stronger populist views tend to trust the news media less.¹¹² Studies show that extreme ideology is positively associated with beliefs in conspiracy theories in Sweden, and in the United States it is also a predictor of lower trust in legacy news media, although those engaged with conspiracy theories may still have an interest in news.¹¹³ Interest in and knowledge about politics are found to influence media trust, and several cross-country studies show that interest in politics is positively associated with trust in news media.¹¹⁴

Research suggests that the news media does not necessarily exacerbate mis- and disinformation problems. A study in 2023 in Brazil, India and the United Kingdom investigated the effect of news and platform use on awareness of and belief in Covid-19 ‘misinformation’. This found that news consumption weakened the acquisition of false beliefs depending on the information access mode (online or offline) and the news outlet type.¹¹⁵

The reasons people distrust the news are also varied. Perceived convergence between the interests of journalists and politicians or businesses

¹⁰² Deuze (2014).

¹⁰³ Boyd (2014).

¹⁰⁴ Jakob (2012); Tsifti & Ariely (2014).

¹⁰⁵ Frieden (2014); Pinkleton *et al.* (2012).

¹⁰⁶ See Schranz *et al.* (2018); Toff *et al.* (2021a) with support of the Facebook Journalism Project; Tsifti & Ariely (2014).

¹⁰⁷ Toff *et al.* (2021a) with support as above.

¹⁰⁸ See Hanitzsch *et al.* (2018); Tsifti & Ariely (2014).

¹⁰⁹ Trauthig (2024).

¹¹⁰ Toff *et al.* (2021a) with support as above; Verma *et al.* (2018).

¹¹¹ Eddy (2024). On the origins of this kind of asymmetric ‘propaganda’, see Benkler (2020).

¹¹² Hanitzsch *et al.* (2018); Stroud & Lee (2013); Suiter & Fletcher (2020), funded by Google UK, part of Google News Initiative, the Broadcasting Authority of Ireland and the Faculty of Humanities & Social Sciences, Dublin City University, Ireland.

¹¹³ Krouwel *et al.* (2017); McKernan *et al.* (2023).

¹¹⁴ Hanitzsch *et al.* (2018); Tsifti & Ariely (2014).

¹¹⁵ Altay *et al.* (2023b, p. 1).

and a belief that the powerful push an agenda is one reason, and this is found to be strong among young people and those with lower incomes.¹¹⁶ An interview-based study suggests that suspicion about the neutrality of news media leads media users to doubt the media, who to trust and what to believe.¹¹⁷ How the media industry reports news also influences trust, with perceived accuracy, impartiality, expertise and integrity shaping the perceived quality of news and the level of trust.¹¹⁸ People's subjective perceptions of accuracy influence trust. For example, when people with direct experience of an event believe there is a difference between what happened and its reporting, their trust in the media will be impacted.¹¹⁹ Research also shows that rumors that go viral are often more influential than the credibility of a source of information or its factuality – sharing such information is found to be motivated less by the accuracy of information than by 'partisan support, community sentiment, emotional contagion and a taste for the sensational or bizarre'.¹²⁰ However, the operationalization of measures of affect or emotion so far relies on inconsistent definitions.¹²¹

As indicated, the role played by news media in circulating what is now called mis- and disinformation (formerly 'propaganda') long predates the internet. The affordances – that is, the instrumental and social features that result from users' interaction with technology – of platformized media change the distribution of power between the news media and its audiences. This raises many questions about the role of algorithms in shaping public beliefs and behaviors.¹²² As digital platforms infiltrate people's lives, this is seen as constituting an epistemic crisis that threatens democracy.¹²³ To understand this, research seeks to measure the effects of audience

exposure to mis- or disinformation to explain the effects of news media on people's attitudes and behaviors. Other research examines the information 'crisis' by studying reciprocal relationships between the content provided by the news media, the roles of changing technologies and the broader political, social, cultural and economic context in which news media operate.¹²⁴

In summary, the problems associated with mis- and disinformation are researched across multiple disciplines. Some studies treat conspiracy theories and pseudoscience as mis- and disinformation, while others do not.¹²⁵ Inconsistent results of research on the effects of mis- and disinformation on democracy, trust and political institutions are partly attributable to different conceptualizations and definitions and to siloed disciplinary research streams. In some cases, the reliability of research findings is questioned. For example, in late 2024 it was revealed that some study results should be questioned, with researchers arguing that a study on the impacts of mis- and disinformation had been influenced by a temporary change in Meta's news algorithm so that it appeared to feed largely reliable sources of trustworthy news to users in contrast to the less rigorous standard algorithm that was normally used. It was argued that this change in the algorithm was not taken into account.¹²⁶ Meta, however, insisted that it had informed the researchers of the change.

4.3 NEWS MEDIA USE, NEWS AVOIDANCE AND RESILIENCE

Numerous factors influence people's media use, whether they try to avoid the news and whether they are likely to be resilient to mis- and disinformation. Where news media are diverse and

¹¹⁶ Newman & Fletcher (2017), supported by Google and the Digital News Initiative.

¹¹⁷ Toff & Nielsen (2018), supported by Google UK as part of the Digital News Initiative.

¹¹⁸ Kantar Media (2016), an international market research company based in London and supported by Google's Digital News Initiative.

¹¹⁹ Livio & Cohen (2018).

¹²⁰ Rodríguez-Ferrándiz (2023, p. 15), supported by the Ministry of Science and Innovation (MCIN) (Ministerio de Ciencia e Innovación), Spain and the European Commission.

¹²¹ Altay *et al.* (2023a), citing Rogers (2020), supported in part by the Connecting Europe Facility and Reboot Foundation. Bakker & Lelkes (2024); Wardle (2023), supported by the Dutch Research Council (NWO, Nederlandse Organisatie voor Wetenschappelijk).

¹²² See Benkler (2020); Guess *et al.* (2023a), supported by Meta, which did not have the right to prepublication approval.

¹²³ For literature reviews, see Ross Arguedas *et al.* (2022a); Tucker *et al.* (2018).

¹²⁴ See Schünemann (2022), for a discussion of research in a socio-technical tradition.

¹²⁵ See the definition of mis- and disinformation in Section 3, Chapter 1.

¹²⁶ See Bagchi *et al.* (2024), supported by a data-sharing agreement with Meta (with no involvement of Meta in the study) and in part by the Knight Foundation and Swiss National Science Foundation (SNSF); for a critique, see Guess *et al.* (2023b), supported by Meta (with no right to prepublication approval) as well as the Democracy Fund, Hopewell Fund, Guggenheim Foundation, John S. and James L. Knight Foundation, Charles Koch Foundation, Hewlett Foundation and Alfred P. Sloan Foundation.

(relatively) free from state coercion, a high choice news media environment presents opportunities to study which audiences consume what type of news media and their consumption patterns to explain political participation.¹²⁷ People are found to have different ‘media repertoires’, that is, engagement to varying degrees with media sources and types of content such as entertainment, political information, regional or national news.¹²⁸

Research finds that age is a strong predictor of media use repertoires, with consistent evidence that older people tend to use more legacy media (i.e., watching television more than other age groups) with younger people getting their news from social media.¹²⁹ The less educated are more likely to access news from television. People who listen to podcasts tend to be more educated, and there is evidence that short videos are becoming a more common source of news, especially for younger people, although this varies by country.¹³⁰ More educated men are more likely to consume news from legacy media.¹³¹

These varied media repertoires are associated with different forms of political participation. In the United States a healthier information ecosystem has historically been associated with a public who consume ‘hard news’.¹³² ‘Soft news’ and social media tend to be regarded as less noble in the research literature, but are confirmed as being important in shaping people’s engagement in politics.¹³³ Despite inconclusive results on whether there is a direct association between political knowledge and incidental news exposure, incidental news exposure is found to lead to reflections on politics that can increase people’s knowledge. In a world where

a humorist can live-cast a conversation with an actor about politics, the distinction between hard and soft news is fragile, and the news media and audiences play a role in influencing the topics that humorists address.

Some people tend to actively avoid certain sources of information. This is problematic in the presence of polarization and partisanship, especially if news selectivity leads to more extreme political positions.¹³⁴ Selective news exposure is influenced by several factors, such as confidence in one’s judgments and political knowledge, or the degree of belonging to a homogeneous social group.¹³⁵ Research shows that while both political knowledge and interest are predictors of news usage, knowledge is a stronger predictor of whether people are more likely to seek out news stories rather than avoid them.¹³⁶ Affective or emotional engagement plays an important role in people’s news usage or avoidance.¹³⁷ It is also important to undertake smaller qualitative studies of everyday news use to reveal the importance of social and cultural dynamics that influence motivations to engage with the news and to share false information on social media and chat apps.¹³⁸

4.3.1 News Media Avoidance

Studies of news avoidance – people who voluntarily or involuntarily consume very little or no news at all – indicate that this is present to varying degrees around the world.¹³⁹ A study published in 2024 with evidence from 46 countries (with data from 2015 to 2022) shows that the number of people claiming not to participate in any news increased by 19%, with this pattern being present in most countries and for most types of news.¹⁴⁰

¹²⁷ Chadwick (2017); Prior (2005).

¹²⁸ Castro *et al.* (2022).

¹²⁹ Castro *et al.* (2022); Kim (2016); Strömbäck *et al.* (2018); Taneja *et al.* (2012), part-funded by Sequent Partners, a marketing consultant, US.

¹³⁰ Aalberg *et al.* (2013); Newman *et al.* (2024), supported by a range of public and private funders including BBC News, Ofcom and Google News Initiative.

¹³¹ Castro *et al.* (2022); Strömbäck *et al.* (2018).

¹³² Schudson (1978).

¹³³ Castro *et al.* (2022); Reinemann *et al.* (2012).

¹³⁴ Buturoiu *et al.* (2023).

¹³⁵ Metzger *et al.* (2020), funded by the John D. and Catherine T. MacArthur Foundation, US; see also Dubois & Blank (2018), supported by Google.

¹³⁶ Lecheler & de Vreese (2017).

¹³⁷ Corbu *et al.* (2021); Zhu *et al.* (2024).

¹³⁸ Tully (2022) demonstrates this in the case of Kenya.

¹³⁹ Skovsgaard & Andersen (2020). For a comprehensive treatment of news avoidance, see Toff *et al.* (2023).

¹⁴⁰ Altay *et al.* (2024), supported by Google News Initiative and the European Commission. The sample based on Reuters Institute’s Digital News Reports overrepresents the Global North; all countries were not present in all years, and in some countries, participation was flat (e.g., Austria, Ireland, France, Japan, Netherlands, Switzerland) or increasing (e.g., Colombia, India, Indonesia, Nigeria, Peru) – but over a shorter time span in recent years.

There are variations in how news avoidance is defined, which means results are not easily compared.¹⁴¹

News media avoidance. A study drawing on the Reuters Institute's *Digital News Report 2023* data found that 'people who selectively avoid news consume almost as much news as those who do not', although this involves a mix of deliberate choices and socially conditioned preferences. The Institute's *Digital News Report 2024* found that 39% of survey respondents said that they 'sometimes' or 'often' avoid the news. This was an increase of 3% over the previous year, and there were more significant increases in Brazil, Finland, Germany and Spain.¹⁴² In France, another survey study found that 94% of people aged 15 and older reported an interest in information, and that they stayed informed daily.¹⁴³

the news.¹⁴⁸ Studies also highlight the fact that people avoid the news when they perceive it to be too pessimistic.¹⁴⁹

News avoidance may be a strategy to protect one's mental health or to avoid information overload (as found in Argentina, Finland, Israel, Japan and the United States).¹⁵⁰ News avoidance can be attributed to a coping mechanism or to a form of protection against the negativity and constant stimulation provoked by contemporary information ecosystems.¹⁵¹ People also avoid the news because they report that it is irrelevant to them, or they believe the news is not trustworthy or that it is too commercial.¹⁵² In Argentina, people were found to avoid the news because they regarded its information ecosystem as corrupt, while in Japan, they were more likely to try to avoid controversy and disagreement.¹⁵³ Notwithstanding these differences reported by individuals, the few studies that look at structural factors find that greater press freedom and political freedom and stability are negatively correlated with news avoidance.¹⁵⁴

There is no normative answer as to how much news people should consume. However, news avoidance is problematic if it isolates people from daily political discussions and political decision-making.

4.3.2 Resilience to Mis- and Disinformation

Declining trust in the news media is associated with declining trust in institutions generally, and there are fears that this is contributing to democratic backsliding. How resilient are people to mis- and disinformation?

Numerous factors are associated with higher levels of news avoidance. Studies point to significantly lower consumption or higher avoidance of news by women.¹⁴⁴ The young are more likely to avoid the news.¹⁴⁵ The more educated access more news than the less educated, according to a longitudinal study in Norway.¹⁴⁶ A longitudinal study in Sweden showed that political interest plays an increasing role in the consumption of news over time – political interests are found to be a determinant of news avoidance, with those with lower interest in politics being more likely to avoid the news.¹⁴⁷ Those with a weaker understanding of the news media ecosystem or who tend to trust the media less are more likely to avoid

¹⁴¹ Bos *et al.* (2016); Castro *et al.* (2022); Strömbäck *et al.* (2018).

¹⁴² Palmer *et al.* (2023, p. 697), supported in part by Google UK as part of the Digital News Initiative; Arcom (2024); Newman *et al.* (2024).

¹⁴³ Arcom (2024).

¹⁴⁴ Toff & Kalogeropoulos (2020); Toff & Palmer (2019), supported by Google UK as part of the Digital News Initiative.

¹⁴⁵ Toff & Kalogeropoulos (2020).

¹⁴⁶ Karlsen *et al.* (2020), supported by the Research Council of Norway.

¹⁴⁷ Edgerly (2022), supported in part by the Walter Jay and Clara Charlotte Damm Fund of the Journal Foundation, US; Strömbäck & Shehata (2019), funded by the Axel and Margaret Ax:son Johnson Foundation, Sweden.

¹⁴⁸ Boukes & Vliegenthart (2017); Edgerly (2022); Toff & Kalogeropoulos (2020).

¹⁴⁹ Aharoni *et al.* (2021); Newman *et al.* (2024). See also Villi *et al.* (2022), supported in part by the Helsingin Sanomat Foundation, Finland.

¹⁵⁰ Aharoni *et al.* (2021); Villi *et al.* (2022), supported in part by the Helsingin Sanomat Foundation, Finland.

¹⁵¹ Ytre-Arne & Moe (2021), funded by the Research Council of Norway; Suiter & Fletcher (2020). Evidence shows that news avoidance grew during the first year of the pandemic.

¹⁵² Aharoni *et al.* (2021); Edgerly (2022).

¹⁵³ Villi *et al.* (2022), a study of Argentina, Finland, Israel, Japan and the United States, supported in part by the Helsingin Sanomat Foundation (Helsingin Sanomain Säätiö), Finland.

¹⁵⁴ Toff & Kalogeropoulos (2020).

Whether declining news media trust is a ‘crisis’, as is sometimes claimed, depends on the country, and the strong focus on trust is questioned by some scholars.¹⁵⁵ However, the instrumentalization of a ‘lying press’ by far-right political leaders is placing trust in the news media at the center of contemporary preoccupations. Cross-country comparative research sheds light on the factors that seem to make people in some countries more resilient to mis- and disinformation (although much scholarship focuses on the United States and Europe). Distrust in legacy media has been associated with alternative news media consumption, which is a cross-national factor associated with declining resilience to mis- and disinformation.¹⁵⁶ However, research also finds that trust in national news media does not build individual resilience, measured as a willingness to share, like or comment on misinformation (in the case of Canada, France and the United States, but not the United Kingdom). In 2020, the use of PSM (e.g., the BBC, France Télévisions or the Canadian Broadcasting Corporation) was not found to build resilience in the short term.¹⁵⁷ A comparison of 18 Western democracies identified three groups of resilience to mis- and disinformation.¹⁵⁸

Country clusters on resilience. Cluster 1: High resilience to mis- and disinformation – Northern and Western European countries (Austria, Belgium, Denmark, Finland, Germany, Ireland, the Netherlands, Norway, Sweden, Switzerland and the United Kingdom), plus Canada; in 2022, all the countries were seen as news media-supportive, more prone to political consensus, less polarized and less prone to populist communication, and characterized by high levels of media trust, shared media consumption and strong PSM. **Cluster 2:** Southern European

countries (Greece, Italy, Portugal and Spain) characterized by low resilience, with high levels of polarization, populist communication and social media news use, and low levels of trust and shared media consumption. **Cluster 3:** The United States – a low-trust, politicized and fragmented political and media environment.

A follow-up study in 2023 indicated that resilience was partly country-specific and highly dependent on the political and information environments.¹⁵⁹ Focusing on Belgium, France, Germany, Switzerland, the United Kingdom and the United States, chosen for their diversity in terms of resilience factors, resilience was measured as an inclination to ignore or disregard ‘disinformation’ as opposed to engaging with it.

Mis- and disinformation and trust in Chile.

A weak relationship was found between mis- and disinformation and media skepticism in 2017 to 2019. Initial beliefs about factually dubious information were negatively correlated with levels of trust in the news media.¹⁶⁰ Although lower trust in the media was related to higher levels of mis- and disinformation, the strength of this association weakened over time. There was no evidence of a positive feedback loop – the reverse spiral model – between mis- and disinformation and media skepticism.

Apart from a limited number of cross-national indicators of resilience (i.e., heavy social media use, the use of alternative media and populist party support), other variables, such as extreme ideology, populist support, age, level of education and

¹⁵⁵ Jakobsson & Stiernstedt (2023), supported by the Swedish Research Council (Vetenskapsrådet), Sweden.

¹⁵⁶ Humprecht *et al.* (2023), supported by the Swiss National Science Foundation (SNSF) and Research Foundation – Flanders.

¹⁵⁷ Boulianne *et al.* (2022).

¹⁵⁸ Based on seven dimensions of media use to create indices of populism, polarization, media trust, shared media, strength of PSB, social media and market size; see Humprecht *et al.* (2020), supported by the Swiss National Science Foundation (SNSF) and Research Foundation – Flanders. Experience of the United Kingdom during the Brexit campaign and since, when polarization increased and mis- and disinformation flourished, indicates that risks exist in this highly resilient cluster.

¹⁵⁹ Humprecht *et al.* (2023), supported by the Swiss National Science Foundation (SNSF) and Research Foundation – Flanders.

¹⁶⁰ Valenzuela *et al.* (2022), a three-wave panel study supported by the National Agency for Research and Development (ANID, Agencia Nacional de Investigación y Desarrollo), Chile.

gender, varied by country. The *contextual nature of resilience* was validated by another, which focused on awareness of, exposure to, and sharing of misinformation.¹⁶¹ Despite inconsistent results, these studies do suggest that engagement with online news media is among one of the most important factors influencing societal resilience. This is in line with studies that conclude that the corrosive effects of mis- and disinformation on attitudes toward the news media are less serious than often assumed.¹⁶²

4.3.3 Weaponization of Online Information

Politicians increasingly seem to be able to lie without negative consequences, and disparate actors – some political elites or digital platform owners – are claiming hegemony over what counts as ‘truthful’ interpretations of reality.¹⁶³

In the late 1960s Hannah Arendt discussed whether it is always necessary to tell the truth, distinguishing between factual truths (facts, events) and rational truths (e.g., mathematical, scientific and philosophical truths) in political debates in plural societies, exploring the disturbing consequences of denying, mystifying or replacing truths with the opinions of political actors.¹⁶⁴ In a ‘post-truth politics’ era, authoritative figures center political communications around the strategic denial of verifiable facts.¹⁶⁵ In this way, information is weaponized, contributing to a democratic crisis. This is especially so when minorities are singled out via social media accounts for receiving divisive and manipulative content.¹⁶⁶ The weaponization of information in political contexts, including elections, is a major concern in many countries. The political use of social media and data to target communications directly at followers in unethical (and sometimes illegal) ways to influence election outcomes is at the core of debates about the harms associated with mis- and disinformation.¹⁶⁷

Cambridge Analytica, the political campaign company that operated from 2013 to 2018, sparked outrage as one of the first ‘information operations actors’ to interfere with democratic processes on a grand scale by microtargeting individual voters and spreading disinformation. It was found to have undertaken illegal data gathering in both the United Kingdom and the United States due to its use of some 5,000 data points on voters, which it secured without user consent via This is Your Digital Life, an app hosted on Facebook. More generally, the company’s tools for targeting voters were used to discredit its clients’ political opposition in numerous countries.¹⁶⁸

In the European Union, the Digital Services Act of 2022 requires the largest platforms, including search engines, to address the systemic risk of ‘negative effects on civic discourse and electoral processes’ associated with their services, but clear benchmarks need to be established.¹⁶⁹ Encompassing more countries, the Council of Europe 2022 recommendation on media coverage of election campaigns states that the ‘the algorithms used by public and private actors to rank and display political advertising and electoral communication material, and those used in content moderation practices, should be transparent and verifiable, especially regarding potential bias and inaccuracies of the systems used’. Platforms are recommended to ‘act against misrepresentation and the intentional spread of political disinformation, while ensuring full respect for the rule of law and human rights standards ... notably the right to freedom of expression’.¹⁷⁰

The *Electoral Integrity Global Report 2024* indicated that the top five countries in terms of election integrity were Czechia, Finland, the Netherlands, New Zealand and Switzerland, and the bottom five were Cambodia, Egypt, Madagascar, Turkmenistan and Zimbabwe. However, from 2012 to 2023 there

¹⁶¹ Boulianne *et al.* (2022).

¹⁶² Allen *et al.* (2020), supported by the Nathan Cummings Foundation, US.

¹⁶³ Hofmann (2024).

¹⁶⁴ Arendt (1968).

¹⁶⁵ Giusti & Piras (2021); Lockie (2017); Merenda (2021).

¹⁶⁶ Freelon *et al.* (2022); Freelon & Wells (2020); Park *et al.* (2023).

¹⁶⁷ Elishar-Malka *et al.* (2020).

¹⁶⁸ Briant (2021); Dowling (2022), supported by the Department of Defence, Australia.

¹⁶⁹ EC (2022c, Article 34(c)); and see Broughton Micova & Schnurr (2024).

¹⁷⁰ Council of Europe (2022, para. 4.2).

were no statistically significant increases or decreases on electoral integrity indices across 586 elections in 170 countries. The survey questions asked about the role of the media, and whether mis- or disinformation was spread on social media as one of several indicators.¹⁷¹

It is known that mis- and disinformation can be advantageous to political figures whose supporters share this content (e.g., Donald Trump and QAnon conspiracy theorists). The Trump Administration and Fox News facilitated and co-produced persistent mis- and disinformation during the coronavirus crisis.¹⁷² Research has demonstrated that Trump's supporters influence the dynamics of top 'fake news' spreaders.¹⁷³

Analysis of X/Twitter news activity suggests that 'fake' and extremely biased news have distinct diffusion mechanisms compared to center- and left-leaning news. In Brazil, a mixed-methods study found that mis- and disinformation tended to circulate more on political pages/groups aligned with the far right and former Brazilian President Jair Bolsonaro, and on religious and conspiracy theory pages/groups and alternative (hyper-partisan) media, whereas fact-checked news circulated more on leftist pages/groups.¹⁷⁴ Another study documented the spread of conspiracy narratives in Brazil about George Soros, providing evidence of cross-platform dissemination.¹⁷⁵ However, in the case of electoral 'misinformation' during the 2022 Brazilian presidential election, it was found that professionally produced news from legacy news organizations played a key role in curbing misinformation, and despite misinformation spreading on digital platforms, there were either no or very small effects between platform use to source news and beliefs in electoral misinformation.¹⁷⁶

Mis- and disinformation practices are consistently associated with far-right political movements and politicians in Brazil, Germany, Portugal and the United Kingdom.¹⁷⁷ Studies of migrant-related mis- and disinformation in the European Union reveal a mix of state-driven activities linked to Russia and domestic far-right actors.¹⁷⁸

Legacy media weaponizing information

in autocracies. This varies by social and political context. Right-wing activists are found to spread their messages by manipulating legacy media and working strategically with partisan media, and there is less research on the magnitude and character of left-wing mis- and disinformation activities. Illiberal political leaders adopt mis- and disinformation as a tool for gaining support and reducing resistance without resorting to terror – by securing formal ownership of the media or informal control, as in the case of 'information autocrats', such as Lee Kuan Yew in Singapore, Augusto Pinochet in Chile, Vladimir Putin in Russia, Alberto Fujimori in Peru and Mahathir Mohamad in Malaysia, and when there is convergence between parts of the media and far right political parties (e.g., the far right party, Vox, in Spain). In the Middle East, the Arab news media is subject to persistent repression by authoritarian governments, with evidence of media pushing mis- and disinformation. China and Russia deploy state-owned media outlets such as Russia Today (RT) and China Central Television (CCTV).¹⁷⁹

¹⁷¹ Garnett *et al.* (2024), supported by the Social Sciences and Humanities Research Council (SSHRC) of Canada, based on electoral authority expert surveys in 42 countries.

¹⁷² Jeppesen *et al.* (2022); Yang & Bennett (2022). For a literature review on how social media profiles are used to manipulate public opinion based on 369 articles, see Santini *et al.* (2018).

¹⁷³ Boulianne *et al.* (2022), part-funded by the Digital Citizenship Initiative of the Department of Canadian Heritage; Bovet & Makse (2019); Pérez-Curiel *et al.* (2021).

¹⁷⁴ Recuero *et al.* (2022).

¹⁷⁵ On the amplification of news content using bots in Brazil, see Santini *et al.* (2022); Santini *et al.* (2020), supported by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), Brazil.

¹⁷⁶ Mont'Alverne *et al.* (2024), funded by the Meta Journalism Project.

¹⁷⁷ Baptista & Gradim (2022); Buarque & Zavershinskaia (2022); Daniels (2018); Freelon *et al.* (2022); Recuero *et al.* (2020); Wojczewski (2022).

¹⁷⁸ Chavalarias (2024); Szakacs & Bogner (2021).

¹⁷⁹ Compiled from Douai (2019), IEMed, a think tank Barcelona, Spain; Freelon & Wells (2020); Guriev & Treisman (2019); Labio-Bernal & Manzano-Zambruno (2023).

In certain African countries, mis- and disinformation are shown to be connected to colonial legacies of mis- and disinformation and propaganda rather than to the rise of the far right.¹⁸⁰ In the political landscapes of countries such as Kenya, Nigeria and Zimbabwe, the spread of false information is especially complex when it is polarized and ethnically charged.¹⁸¹ All sides of the political spectrum, as well as legacy and social media, participate in mis- and disinformation.¹⁸² It is important to note that while mis- and disinformation or 'fake news' is seen as a novel scholarly topic today, false news as a phenomenon in Africa and the Middle East pre-dates the era of online news. Journalists have always had to learn to treat journalism as a contested area, vulnerable to manipulation by governments and powerful social elites.¹⁸³ However, recent developments have provided new opportunities for governments to restrict freedom of expression on social media.

State-sponsored mis- and disinformation campaigns are common in multiple political systems.¹⁸⁴ For example, tools, capacities, strategies and resources for computational propaganda have been identified in 81 countries, with private firms engaged in manipulation campaigns and practices of harassment against fact-checkers and those reporting on information operations.¹⁸⁵ Many illiberal leaders are preserving a democratic facade while controlling the information space – acting as 'informational autocrats'.¹⁸⁶

Anti-Western propaganda is characteristic of Russian influence operations and information warfare aimed at undermining trust in NATO, the European Union and domestic governments, by

interfering in elections and undermining democratic processes.¹⁸⁷ Actors, especially in Russia and China, have been connected to mis- and disinformation campaigns in both the Global North and Global Majority World.¹⁸⁸ Russian propaganda pushes mis- and disinformation narratives with the aim of winning a significant share of media audiences in countries such as Argentina, Bolivia, Colombia and Mexico.¹⁸⁹ China has been shown to be waging a state-sponsored mis- and disinformation campaign against a United States-led international system aiming to suppress internal and external criticism, amplify its prestige and favorably influence foreign policy actors, and it engages in information warfare to support its military strategy (e.g., on the issue of Taiwan).¹⁹⁰

An analysis of the activities of troll factories under the control of the Russian Internet Research Agency (IRA) from 2017 to 2019 in the United States found that the scale of the IRA's troll factories was 'industrial – mass produced from a system of interchangeable parts, where each class of part fulfilled a specialized function'.¹⁹¹ This highlights how, rather than posting tweets in support of one party, IRA trolls tweet divisive messages, sometimes targeting mainstream Republicans and, at other times, mainstream Democrats – tactics consistent with the aim of sowing mistrust and doubt in the election process.¹⁹²

Research during the United States on the 2020 presidential election examined the promotion by pro-Kremlin media (the channel RT, in particular) on Facebook, and how content curation algorithms affected its distribution,¹⁹³ finding that the Facebook News Feed algorithm (which Facebook describes as aiming to expose users to reputable

¹⁸⁰ Mudde (2019).

¹⁸¹ Mare *et al.* (2019).

¹⁸² Lunga & Mthembu (2019); Ncube (2019); Wasserman (2020a).

¹⁸³ Mutsaers & Bebawi (2019).

¹⁸⁴ La Cour (2020).

¹⁸⁵ Bradshaw *et al.* (2021), supported in part by the European Research Council (ERC), Adessium Foundation, Civitates Initiative, Ford Foundation, Hewlett Foundation, Luminate, Newmark Philanthropies and Open Society Foundations.

¹⁸⁶ Guriev & Treisman (2019).

¹⁸⁷ Akimenko & Giles (2020); Alieva *et al.* (2022); Beskow & Carley (2020); Lemke & Habegger (2022); Morkūnas (2023); Robbins (2020); Zhang *et al.* (2021), supported by the Knight Foundation and Office of Naval Research, US.

¹⁸⁸ Chaguaceda *et al.* (2023); Sleibi (2023); Znojek (2020).

¹⁸⁹ Chaguaceda *et al.* (2023).

¹⁹⁰ Cheng (2016); Curtis (2021); Hung & Hung (2022), supported by the Tzu-Chieh Hung Ministry of Science and Technology, Taiwan.

¹⁹¹ Linvill & Warren (2020, p. 463).

¹⁹² Linvill & Warren (2019), supported by the Charles Koch Foundation, US.

¹⁹³ Kuznetsova & Makhortych (2023, p. 22).

information¹⁹⁴) still make it possible for pro-Kremlin media to propagate these messages, indicating that the Facebook content curation is vulnerable to manipulation of ‘likes’ to enhance message flows.

The weaponization of discourses using concepts such as ‘fake news’ is typically invoked in competitions for power, and is used to discredit, attack and delegitimize political opponents.¹⁹⁵ Legitimate news media are targeted by ‘fake news’ labeling,¹⁹⁶ reducing the perceived credibility of authentic media content, although research suggests that this may not affect people’s policy preferences.¹⁹⁷ The weaponization of information is

coincident with the explosion of technologies that help to make mis- or disinformation part of a flux of overabundant information.¹⁹⁸

Although search engines play a role in promoting mis- and disinformation, much attention focuses on the role social media plays in the creation, distribution and monetization of this online content because it is sometimes outperforming legacy media as a source of news.¹⁹⁹ Research on the large social media platforms (Facebook, X/ Twitter, Instagram, TikTok), video-sharing platforms (YouTube) and main private messaging apps (WhatsApp, Telegram) is summarized in Table 2.2.

Table 2.2

Platform roles in the weaponization of information – Selected Country Examples

Country/region	Example	Platform in question
Argentina (49) Bolivia (27) Colombia (78) Ecuador (65) Peru (28) Spain (127)	Channels of distribution across six Spain and Latin American countries: Facebook was the most-used network to disseminate mis- and disinformation (32.9%), followed by hoaxes disseminated in two or more networks (31.9%), WhatsApp (21%), Twitter (5.7%), email or SMS with (5.4%) and YouTube (3.0%).	Multiple social media platforms; bibliometric study of 371 examples of mis- and disinformation about the Covid-19 pandemic (March-May 2020).
Australia (2) England (1) Spain (1) United States (8)	Weaponized crowdfunding by actors to amplify and sustain the spread of their grievances. Crowdfunding platforms offer comprehensive tools that facilitate easy sharing and propagation of campaign messages across various platforms.	GoFundMe, a set of election fraud and 5G-themed campaigns on the crowdfunding platform GoFundMe.
Canada	Prevalence of misinformation surrounding Covid-19 on Twitter, compared to Canadian news media: social media exposure was associated with more misperceptions and less social distancing compliance.	X/Twitter; all articles published on 19 Canadian news sites.
India	WhatsApp as a tool for political communication used by political parties in India: for mobilization, coordination and reaching out to voters; political propaganda and disinformation were pushed on WhatsApp in the form of ‘news’.	WhatsApp.
United States	Evidence from the 2016 presidential election on the virality of political fake news: posts favoring Trump were shared 30 million times on Facebook, while those favoring Clinton were shared 8 million times.	Facebook/Meta.
United States	Fueling civil disobedience in democracy: WhatsApp news is negatively associated with political knowledge and positively with illegal protest.	WhatsApp.
United States	False news stories (2006-17) diffused significantly further, faster, deeper and more broadly than the ‘truth’. This effect was stronger for political news than for other topics.	X/Twitter.

¹⁹⁴ Brown & Levin (2020), a Meta blog post.

¹⁹⁵ Farkas & Schou (2018).

¹⁹⁶ Tong *et al.* (2020).

¹⁹⁷ Hameleers & Marquart (2023).

¹⁹⁸ Bargaoanu & Radu (2018).

¹⁹⁹ Newman *et al.* (2023), supported by a range of public and private funders including BBC News, Ofcom and Google News Initiative; Aimeur *et al.* (2023); Wakefield (2016), supported by the Natural Sciences and Engineering Research Council, Canada.

Country/region	Example	Platform in question
United States	Amplifying climate mis- and disinformation: showing that posts linking to content from 10 ‘superpolluter’ publishers, totaling 186 million followers on Facebook, accounted for up to 69% of Facebook interactions with climate denial content.	Facebook/Meta. A sample of 6,983 climate denial articles were published between 12 October 2020 and 1 October 2021.
Zimbabwe	Digital propaganda ‘battles’, where political gladiators used mis- and disinformation, hate speech and mudslinging as weapons.	X/Twitter.
N/A	YouTube facilitates access to problematic content (sometimes with mixed results). Nine of the studies demonstrated support for the creation of filter bubble effects.	YouTube’s systematic review of 23 studies published between 2013 and 2021.

Source: Collated from various sources indicated below.²⁰⁰ Note: numbers indicate the incidents covered by the cited sources.

Sponsored content and the absence of transparent political advertising rules play a major role in weaponizing information. The amplification of mis- and disinformation uses the same tools that are the backbone of online advertising (e.g., precision advertising, algorithmic advertising, data-driven behavioral segmentation, ‘psychographics profiling’, computational profiling, computational persuasion).²⁰¹ In the United States, consumption of legacy media sources is found to be associated with more accurate beliefs about health-related topics and consumption of non-partisan, liberal media instead of conservative partisan media, and there was evidence of a smaller inclination to access ‘fake news’ websites.²⁰² These developments are examined in multiple studies that aim to establish whether and to what extent (under what conditions) exposure to mis- and disinformation is causing changes in public opinion and leading to polarization.

4.4 PUBLIC OPINION AND POLARIZATION

Research on causal relationships between mis- and disinformation and ‘filter bubbles’ or ‘echo chambers’²⁰³ and polarized public opinion yields ambiguous or at least contested results.²⁰⁴

Intensive research is underway on these questions, especially since 2014, when social media platforms – Facebook and Twitter (now X) – added ‘like’ and ‘retweet’ buttons to their sites. Whether social media increase political polarization has been a source of inquiry since at least 2017. Some studies show how exposure to content on social media can increase affective or emotional polarization, while others show limited and asymmetrical effects of social media use on attitudes towards people with diverging views.

For example, in 2020, a study in France, the United Kingdom and the United States found no evidence that online social media ‘explain support for right-wing populist candidates and parties’. It did find that ‘offline discussion with those who are similar in race, ethnicity and class positively correlates with support for populist candidates and parties in the United Kingdom and France’.²⁰⁵ In the same time frame, a study in the United States found a ‘substantial amount of overlap (51%) in the ideological distributions of accounts followed by users on opposite ends of the political spectrum’.²⁰⁶ However, in 2022, a study on the role of social media platforms in contributing to

²⁰⁰ Allcott & Gentzkow (2017); Center for Countering Digital Hate (2021); Chibwe (2020); Elmer & Ward-Kimola (2023); Gutiérrez-Coba *et al.* (2020), supported by the Department of Canadian Heritage; Farooq (2018); Gil de Zúñiga & Goyanes (2023), supported by the Spanish National Research Council (CSIC, Consejo Superior de Investigaciones Científicas); Vosoughi *et al.* (2018), supported by Twitter; Bridgman *et al.* (2020), supported by the Department of Canadian Heritage Digital Citizens Initiative.

²⁰¹ Bargaoanu & Radu (2018); Cano-Orón *et al.* (2021); Szczepkowski & Szczepkowski (2021).

²⁰² Jamieson & Albarracín (2020), part-supported by National Institutes of Health (NIH) grants, US; Guess *et al.* (2019), supported by the National Science Foundation (NSF), US.

²⁰³ Pariser (2011, p. 9): ‘a unique universe of information for each of us ... which fundamentally alters the way we encounter ideas and information’ enabled by algorithmic prediction engines; see also Sunstein (2007). Jamieson & Cappella (2008, p. 76) define an echo chamber as ‘a bounded, enclosed media space that has the potential to both magnify the messages delivered within it and insulate them from rebuttal’.

²⁰⁴ Haidt & Bail (2024) provide a review of studies of whether social media: makes people angry or affectively polarized; creates echo chambers; amplifies posts that are emotional, inflammatory or false; increases the probability of violence. Most cited studies are experimental or quasi-experimental and undertaken in the United States and Western democracies. For a systematic analysis of susceptibility to online misinformation in the United States, see Sultan *et al.* (2024), funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) and Volkswagen Foundation (Volkswagen Stiftung).

²⁰⁵ Boulianne *et al.* (2020, p. 683), supported by the Audencia Foundation, France.

²⁰⁶ Eady *et al.* (2019, p. 1), supported by the National Science Foundation (NSF), Knight Foundation and Rita Allen Foundation, US.

radicalization and violent extremism found no effect of echo chambers.²⁰⁷ Another claimed that research in this area overestimates the impact of digital technologies in explaining social and political developments.²⁰⁸ A 2024 study, again in the United States, concluded that AI tools (large language models, in this case GPT-3) in the hands of those launching foreign covert propaganda campaigns can be highly persuasive, as measured by people's agreement with claims made.²⁰⁹ Yet a study of claims about filter bubbles was challenged in another study of public opinion in the United States, which found that social media use had led to less polarization as judged by partisanship (in this case, vaccine hesitancy), while use of legacy media made people more polarized.²¹⁰

Studies of single platforms at a single point in time also indicate, for example, that in 2021, Twitter's personalization algorithm was amplifying tweets in Canada, France, Germany, Japan, Spain, the United Kingdom and the United States, but did not 'support the hypothesis that algorithmic personalization amplifies extreme ideologies more than mainstream political voices'. It did show that the political right experiences higher amplification compared to the political left overall.²¹¹ The study was unable to identify precise causal mechanisms for the variations among countries.

to like-minded content did not reduce polarization, although it did decrease exposure to uncivil language while increasing exposure to cross-cutting sources.²¹² A second study investigated whether Facebook enables ideological segregation in political news consumption, finding that conservatives were more segregated than liberals, and that disinformation circulated mostly in an isolated conservative space. This study found that 'ideological segregation is high and increases as we shift from potential exposure to actual exposure to engagement'.²¹³ A third study examined the effect of Facebook's news feed algorithm by sorting posts chronologically, finding chronological filtering did not affect levels of issue polarization and affective polarization, although it could increase exposure to untrustworthy content and content from moderate voices, as well as decrease exposure to uncivil content and the time spent on the platform.²¹⁴ A fourth study looked at the effects of reshared content, finding that removing reshares from the platform could reduce exposure to untrustworthy content, but this did not affect political polarization.²¹⁵

Social media and polarization. A team of researchers collaborated with Meta to investigate questions about social media effects on politics using large-scale experiments during the United States 2020 presidential election. One study looked at the effects of echo chambers on polarization, finding that reducing exposure

These four studies indicate that personalization systems and interaction with like-minded content can influence consumption of content from untrustworthy sources and exposure to incivility, but they do not demonstrate clear effects of social media on polarization. This research needs to be assessed in the light of the fact that it was conducted over a relatively short time, focused on one country (where polarization has increased

²⁰⁷ Gunton (2022).

²⁰⁸ Talamanca & Arfini (2022), supported in part by the Ministry of University and Research (MUR, Ministero dell'Università e della Ricerca), Italy.

²⁰⁹ Goldstein *et al.* (2024); OpenAI provided access to GPT-3 via an academic access program.

²¹⁰ Jones-Jang & Chung (2024).

²¹¹ Huszár *et al.* (2022, p. 4). Several authors were employed by, affiliated with, or had a financial interest in X/Twitter at the time of the study.

²¹² Nyhan *et al.* (2023), supported by the Facebook Open Research and Transparency (FORT) team and by foundations and universities; some authors were employed by Meta.

²¹³ González-Bailón *et al.* (2023, p. 392). The Facebook Open Research and Transparency (FORT) team provided support for the project; some authors worked for Meta; funding by Meta, Democracy Fund, Hopewell Fund, Guggenheim Foundation, John S. and James L. Knight Foundation, Charles Koch Foundation, Hewlett Foundation and Alfred P. Sloan Foundation. Based on aggregated data for 208 million Facebook users in the United States.

²¹⁴ Guess *et al.* (2023a), supported by Meta.

²¹⁵ Guess *et al.* (2023b), supported by Meta.

in recent years), and on a single platform. Steps were undertaken to ensure the reliability of the findings and to limit Meta's influence on the results, but reliance on one platform company's data can potentially bias the results.²¹⁶

These and other studies demonstrate how hard it is to pin down clear causal relationships between mis- and disinformation, the news media, the role of algorithms and platformization in changes in polarization and political participation.²¹⁷ This may partly be explained by problems of access to platform data, overreliance on controlled experiments and the scarcity of applications of field test methodologies that seek to confirm hypotheses about the causal effects of mis- and disinformation.²¹⁸ The difficulty in clarifying effects of 'social media' on polarization is due to the entanglement of content, personalization systems and social relations, all of which contribute to attitudes towards mis- and disinformation and to political participation. Studies on polarization are resource-intensive, and researchers must generally collect data through platform application programming interfaces (APIs) that may allow limited access.

Monitoring mis- and disinformation, transparency and CrowdTangle.

CrowdTangle was especially valued for its use in mis- and disinformation monitoring. The tool was purchased by Meta in 2016 and allowed researchers, journalists and fact-checkers to explore public content posted on multiple social media platforms, including Facebook, X/ Twitter, Instagram and Reddit. In August 2024, Meta announced that it was shutting the tool down, explaining that this was due to data access changes required by the European

Union's Digital Services Act. The company announced Meta Content Library, which it said would provide the same kinds of services previously available through CrowdTangle. Usage limits, however, mean that many CrowdTangle users are not allowed access, and this has been criticized in the United States as a political move to censor partisan information close to the presidential election.²¹⁹ Some claim that the Meta Content Library has only '1% of its features' of CrowdTangle, expressing doubts about the adequacy of its replacement.²²⁰ This move has been criticized by the European Commission for reducing platform transparency and access to data.²²¹

Studies of rumors and information online and their role in political campaigns, even when their accuracy is uncertain, find that repeat spreaders can disproportionately influence public opinion, although there is also evidence of perverse effects of efforts to raise awareness around 'deepfakes' being associated with distrust in legitimate information.²²² At the same time, research shows that 'identity propaganda' aimed at amplifying historic differences and perpetuating hegemonic power structures can influence public opinion through its use of 'othering' narratives and its influence on attitudes and behavioral norms.²²³ Other researchers argue that explanations for some of the differences in research results would become clearer if research methodologies took account of a wider range of contextual factors including power relationships, rather than focusing on experimental or even field research studies to identify patterns of individual cognition and behavior.²²⁴ Research also finds that the cost of reaching people with mis- or disinformation, not the cost of creating it, is a bottleneck for those intent on distributing

²¹⁶ 2020 Election Research Project (2020).

²¹⁷ Ecker *et al.* (2024); Robertson *et al.* (2024), supported in part by Google Jigsaw and the Templeton World Charity Foundation.

²¹⁸ Forum on Information and Democracy (2024c).

²¹⁹ Gotfredsen & Dowling (2024).

²²⁰ Bellan (2024).

²²¹ Kroet (2024).

²²² Kennedy *et al.* (2022); Twomey *et al.* (2023); Weismueller *et al.* (2023); Guess *et al.* (2023a), supported by Meta, plus various foundations and universities.

²²³ Reddi *et al.* (2023).

²²⁴ Martínez-Costa *et al.*'s (2023) work builds on theories of self-perception, self-efficacy, confirmation bias, miscalibration, misplacement and mis-estimation from psychology and economics.

mis- and disinformation, and that the evidence for the effectiveness of microtargeting through personalization is, in any case, limited.²²⁵

There is much less research on the role of mis- and disinformation in shaping public opinion in non-Western countries. As in the Global North, there is little consensus in the research community on how to define what counts as ‘fake’ or ‘false news’. It is often the state that determines what is to count as misleading information, and this applies in both autocratic and democratic countries. Especially in countries in the Global Majority World, the ontological and epistemic implications of using language like information ‘disorder’, ‘threat’ or ‘pollution’ can be symptomatic of ethnocentrism that privileges a Western view of how information should be generated and assimilated in liberal democracies, with criticisms focusing on who is doing the labeling, what is being labeled as well as on how it is labeled.²²⁶

It is important to note that research in some countries in the Global Majority World points to positive features of online filter bubbles and echo chambers, which are found to provide some degree of respite from targeted attacks on marginalized groups.

Positive features of filter bubbles and echo chambers. This phenomenon is shown to help protect marginalized groups – e.g., feminists, LGBTQ+ populations, those with disabilities, religious groups or political dissidents – by providing a safe space and possibilities for avoiding political or social repression. For vulnerable populations and disadvantaged

or marginalized groups, the appearance of polarized groups communicating in ‘filter bubbles’ can yield safe spaces to express opinions, and well-conceived algorithms have the potential to enable people to express their ideas and identities without fear of punishment.²²⁷

Differences in news media and information use, experiences of harassment and abuse, invasive data collection and propagation of mis- and disinformation are implicated in fanning ‘the flames of hatred and division in society’.²²⁸ Overall, research suggests that exposure to like-minded political content is one of a number of causes of polarization of public opinion. Some argue that ‘politically partisan online news echo chambers are generally small – much smaller than is typically assumed in public and policy debate’,²²⁹ the claim being that research on the negative impacts of mis- and disinformation exaggerates the harms.²³⁰ One expert interviewed for this report suggested that empirical evidence on polarization indicates that mis- and disinformation and social media algorithms contribute to a small extent. While there is little evidence that most people are influenced by the mis- and disinformation they encounter online, elite cues matter more in terms of impact and influence, and the long-term effects are not well understood.²³¹ Another expert observed that the causal relationship may be that polarization in society generally is itself a cause of people falling for mis- and disinformation.²³² As a further expert pointed out, the availability of platforms has allowed people who could not get into mainstream media to get a name for themselves and cover a different side of the news.²³³

²²⁵ Simon *et al.* (2023).

²²⁶ Banaji & Bhat (2022); Banaji *et al.* (2019); Harsin (2024).

²²⁷ Erickson (2024); Toff *et al.* (2021b).

²²⁸ Bennett & Livingston (2020, p. 20).

²²⁹ Ross Arguedas *et al.* (2022a, p. 17).

²³⁰ Altay & Acerbi (2023, p. 2), supported in part by BBC World Service Trusted News Initiative; Allcott & Gentzkow (2017, p. 211); Karpf (2020); McGonagle *et al.* (2019).

²³¹ Interview with Rasmus Kleis Nielsen, then Director of the Reuters Institute for the Study of Journalism, Professor of Political Communication at the University of Oxford, UK, 12 February 2024.

²³² Interview with Natalia Aruguete, Researcher at the National Scientific and Technical Research Council (CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas), Professor at the National University of Quilmes (Universidad Nacional de Quilmes), Argentina, 13 February 2024.

²³³ Interview with Eugenia Mitchelstein, Associate Professor and Chair in the Department of Social Sciences, University of San Andrés (Universidad de San Andrés), Co-Director at the Center for the Study of Media and Society in Argentina (MESO, Centro de Estudios sobre Medios y Sociedad), 27 February 2024. See also Boczkowski & Mitchelstein (2021, 2022); Mitchelstein & Boczkowski (2023).

5 Strengthening Trust and Resilience to Mis- and Disinformation

In addition, cross-disciplinary and longitudinal research is relatively scarce.²³⁴ In some cases, the negative effects of mis- and disinformation on political outcomes and democracy are assumed at the outset of a study, and in others, the impacts on truth/trust in expertise, institutions and the news media are alluded to, but the role of mis and disinformation is unclear or unspecified.²³⁵

Based on this synthesis of research, filter bubbles and echo chambers are ‘not phenomena purely related to algorithms and what information they present, but to how people react to and interact with information’,²³⁶ and this depends on the locale. It is essential to undertake research that considers individual agency, the market structural conditions (financial and business models) in the news media and platform industries and political ideologies that feed polarization, if polarization phenomena are to be understood.²³⁷

Even if vulnerability to mis- and disinformation and its impact on public opinion varies by context, there is no doubt that powerful actors do jeopardize ‘free and open opinion formation as well as promote the dispersal of communicative power’.²³⁸ Research on the effects of filter bubbles and echo chambers on public opinion and polarization needs to acknowledge that what ‘is unproblematic for one individual can have fatal consequences for another... which consequences are negative and which are positive is always contestable’.²³⁹ In the real world of politics and democracy, governance rules and tools are needed to enable people to resist the negative implications of harmful information, while seeking to protect human rights and uphold the normative goals of the news media – and information ecosystems generally. Counterpower is essential if news media and platform power are to be resisted.

Building trust in the news media is crucial for democracy. When trust is low or unevenly distributed over political divides, there are no easy solutions.²⁴⁰ Governance and policy interventions are discussed in Chapters 6 and 7 with a focus on the potential for the exercise of counterpower to both dominant news media and mis- and disinformation. This section is concerned with how journalists and media organizations try to remedy declining trust and trustworthiness when it occurs.²⁴¹

One strategy is for journalists to be more transparent about their work, reporting factual content and avoiding opinions where possible.²⁴² The benefits of sharing information on the process of writing a news story or sharing sources are illustrated by a cross-country initiative to increase transparency – the annual International Journalism Festival, the biggest free and open-to-the-public media event in Europe.²⁴³ The literature suggests that news media organizations can address declining trust by addressing four aspects of news production (see Figure 2.1).²⁴⁴

²³⁴ Obreja (2023).

²³⁵ Kapantai *et al.* (2021, p. 1303), funded by the European Commission; van der Linden (2023, p. 96), citing Vosoughi *et al.* (2018); Eady *et al.* (2023), supported by the National Science Foundation (NSF) and several other US foundations; Nyhan *et al.* (2023), the Facebook Open Research and Transparency (FORT) team provided substantial support; Guess *et al.* (2023a), supported by Meta as well as a variety of US foundations; Allcott & Gentzkow (2017); LSE Truth, Trust & Technology Commission (2018).

²³⁶ Talamanca & Arfini (2022, p. 19), supported in part by the Ministry of University and Research (MUR, Ministero dell’Università e della Ricerca), Italy.

²³⁷ Pickard (2020a).

²³⁸ Seipp *et al.* (2023b, p. 20).

²³⁹ Geiß *et al.* (2021, p. 683).

²⁴⁰ Skovsgaard & Andersen (2020).

²⁴¹ Kohring & Matthes (2007).

²⁴² Newman & Fletcher (2017), supported by Google UK as part of the Digital News Initiative.

²⁴³ See www.journalismfestival.com/faq.

²⁴⁴ Kohring & Matthes (2007); Prochazka & Schweiger (2019).

Figure 2.1
Building trust in news



Source: Banerjee et al. (2023, p. 4)

What the public expects from news media has been investigated in Brazil, India, the United Kingdom and the United States.²⁴⁵ Common expectations were that news organizations should work towards more transparency, achieving better alignment of editorial coverage with concerns in people's everyday lives and preserving media's independence. There were differences with respect to perceptions of newsroom diversity and concerns about one-sided coverage and initiatives to engage more with audiences. This study indicated that those who trust news are more receptive to initiatives that increase audience engagement, whereas the most distrusting individuals are likely to view all media outlets negatively and are harder to reach.

6 Chapter Summary

This chapter has emphasized that what counts as 'news' is hard to define – it includes legacy media organizations, online news producers, mainstream and 'alternative media' and content produced not only by professional journalists but also by a host of other individuals. The focus is mainly on the news industry that employs professional journalists, but the role of actors who produce mis- and disinformation has also been examined.

The analysis of research on asymmetrical power relations between legacy news media organizations and online news media, big tech-owned platforms and their audiences highlighted how the structure of the news media industry and platform dominance of the advertising market are key factors contributing to what is widely seen as an 'information crisis'. Market structures, ownership arrangements and the financial stability of news media organizations differ among countries, as do the offline material conditions in people's lives. These factors affect trust (or mistrust) in news media, and the circulation of mis- and disinformation.

Healthy information ecosystems depend on a robust public sphere. Mis- and disinformation circulating at scale through legacy and online news media were shown to be incompatible with people's fundamental rights 'to hold opinions without interference' and to 'impart information and ideas through any media'. The analysis yielded a complicated picture of what happens when people cannot tell the difference between accurate and inaccurate – or false – information. Questions about who consumes the news, whether they trust it and whether exposure to content is a principal cause of changes in people's attitudes and behaviors that lead to political polarization were shown to be difficult to answer based on existing empirical research.

The research evidence indicates that a focus on the public's declining news media trust (in some countries) needs to be complemented by research on media organizations' responsibilities

²⁴⁵ Banerjee et al. (2023) – a mix of survey research, in-depth qualitative interviews, focus groups and other techniques.

to demonstrate their trustworthiness, although it is also important to note that people access news even if they distrust it. Declining news media trust in Western societies may partly be a response to rising skepticism about political and other institutions. Increasing levels of distrust may be desirable if this is associated with critical thinking, while recognizing that distrust can also be associated with nativist and racist sentiments. As a media historian notes, ‘trust in institutions is salutary for democracy only to a point. The decline in trust in most institutions that public polling has documented since the 1960s was a decline from what was arguably much too unquestioning a level of trust’.²⁴⁶

Declining trust in news media might be a sign of a more intellectually active public. As emphasized in the context of South Africa, research indicates that it is essential for journalists to engage with questions about what constitutes ‘truth’, ‘accuracy’ and ‘facts’ to avoid being branded as the ‘lying press’.²⁴⁷ This is important in all countries where journalism faces ‘moral panics’ around mis- and disinformation.

The design and methodologies of much research is used for assessing the relationships between news media engagement, people’s attitudes and behaviors and how the news media plays a role in the social and political ordering of societies, especially in their information ecosystems. Many studies aim to establish direct causal links between mis- and disinformation, changes in attitudes and behaviors and political polarization. This research is informed by theories of media effects and is undertaken in experimental or quasi-experimental settings or based on survey respondent self-reporting.

Other studies question whether the search for direct causal effects is appropriate. This research also points to the complexity of relationships that are shaped by encounters with news media and mis- and disinformation. These research traditions emphasize the agency of audiences to interpret the information they encounter. In this context, the

relationship between filter bubbles, echo chambers, political polarization and mis- and disinformation is a reciprocal one that depends as much on conditions online as on conditions in the social, political, cultural and economic environment.

Despite the absence of consensus on the specific causes and consequences of distrust in news media and the rise of mis- and disinformation, this distrust is clearly implicated in harms to individuals and society – especially to marginalized and disadvantaged groups.

The synthesis of research in this chapter shows that:

- Dependence of news media on digital platforms places pressure on journalists and news organizations struggling to adjust to declining advertising revenue (PSM also faces varying levels of financial support). News media challenges, including declining advertising, are due in large part to platform business models and the priority given by digital platforms to monetizing audience engagement. Most news organizations are struggling to maintain financial sustainability and independence, and news deserts are emerging in some regions.
- When news media industry concentration rises, this is found to weaken media pluralism and perceptions of the trustworthiness of news organizations. News organizations’ dependence on digital platforms varies by country, type of news, legacy versus online and strategies to address sub-audience segments, including younger audiences. Avenues are needed to strengthen the bargaining power of, especially smaller, news organizations against the platforms.
- State ownership of news media is viewed positively in some democracies and negatively in authoritarian states, but it is widely acknowledged that a diverse news media industry is essential to support a healthy information ecosystem.

²⁴⁶ Schudson (2022, p. 150).

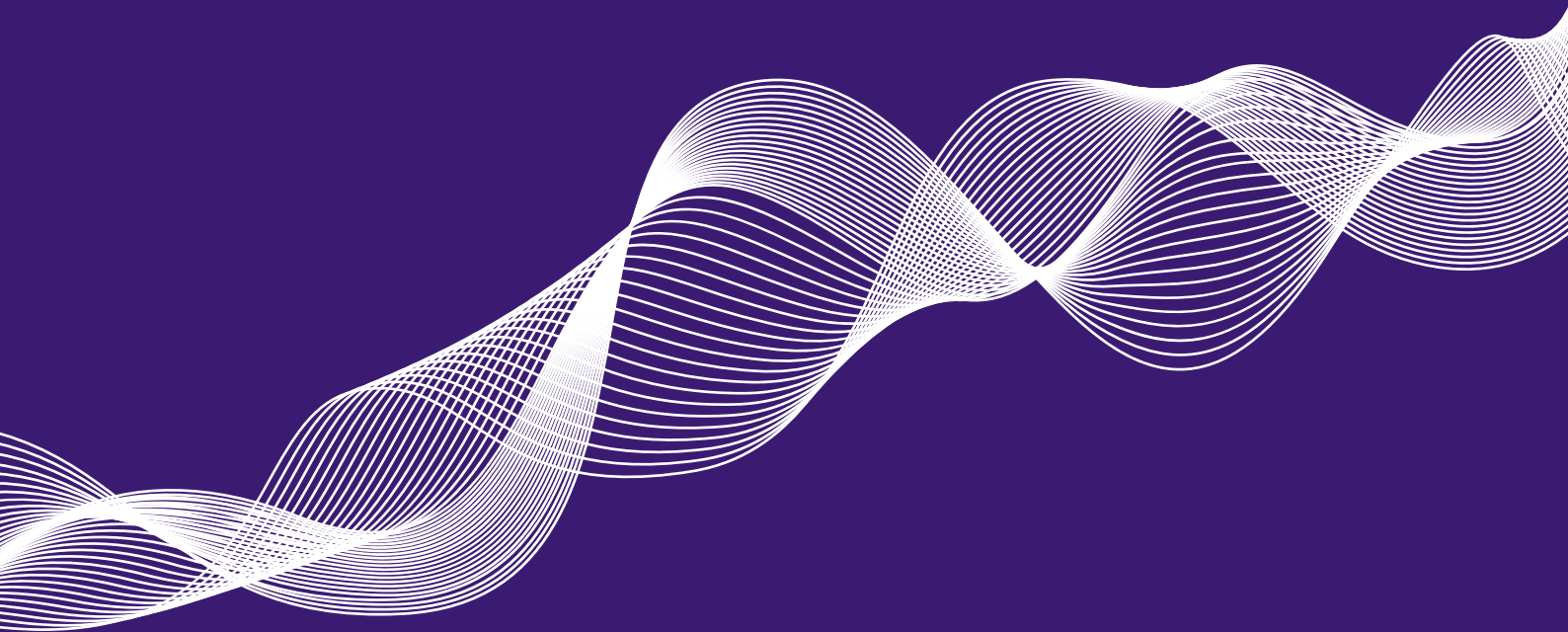
²⁴⁷ Wasserman (2020a).

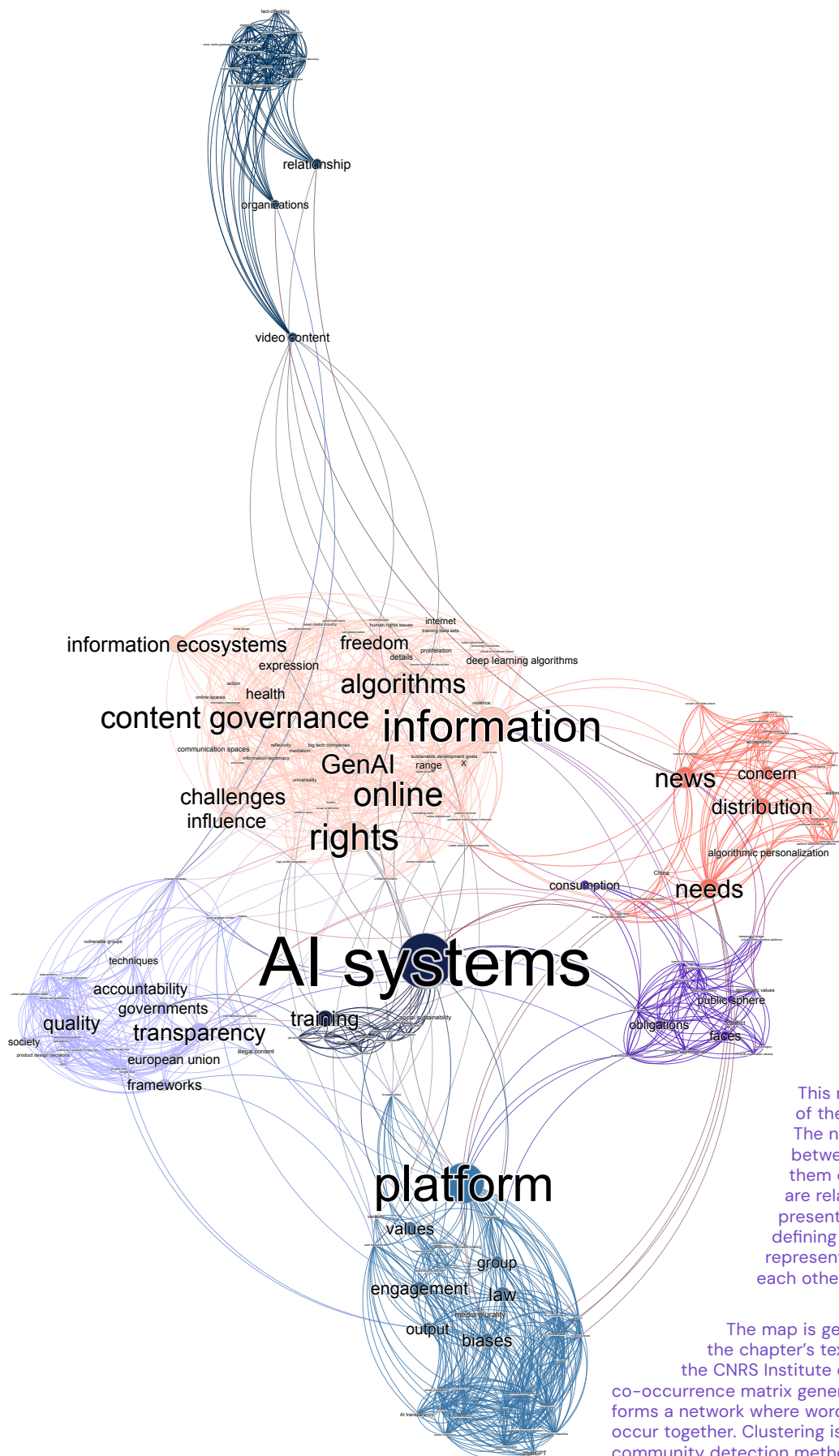
- News media trust depends on variables including age, gender, education, ideology and partisanship and socio-economic status. There are large differences across countries. In democratic and authoritarian countries interest in, and knowledge of, politics influences news media trust; this is linked to interest and participation in politics – and this linkage seems to be becoming stronger over time.
 - News media trust is declining in some countries, stable in others and increasing in yet others. Declining trust in news media (and public authorities) is likely to persist in countries where it is already present.
 - News avoidance is problematic because it isolates people from public life. People report that they avoid the news if it is too pessimistic, to protect their mental health or to cope with information overload.
 - Research on news media consumption confirms that exposure to incidental news grows as social media use increases. Selective news exposure and news avoidance is influenced by factors such as confidence in the ability to discriminate between accurate and false information, political knowledge and whether people belong to homogeneous social groups. People may access news even if they distrust it.
 - The destabilizing effects of mis- and disinformation on political processes are studied mainly in the United States and other Western countries. Countries outside the West are included in some studies, but coverage of the whole of the Global Majority World is patchy.
 - The role of government bodies, ruling political parties and other actors in manipulating information during critical election periods is a concern in many countries. Especially in Global Majority World countries, mis- and disinformation can be due to polarized and ethnically charged politics on all points along the political spectrum. This occurs when legacy and online media engage in the production or circulation of mis- and disinformation.
 - The weaponization of information is often linked to far-right groups doing the bidding of foreign powers, and mis- and disinformation campaigns ramp up in times of conflict.
 - Cognitive biases can lead to overconfidence in abilities to detect mis- or disinformation, and exposure to like-minded political content can be associated with polarization, but partisan online echo chambers are generally found to be smaller than is typically assumed in policy debates.
 - Self-imposed filter bubbles in some contexts can help protect marginalized groups by providing a safe space to express opinions and avoid political or social repression.
- Research is needed:
- To investigate the respective roles of legacy news media, online news media and political actors (as well as other actors) who contribute to mis- and disinformation.
 - To investigate factors contributing to differences between healthy and unhealthy forms of skepticism towards content and information sources.
 - To undertake longitudinal studies with global coverage to assess relationships between changes in media trust and in political polarization and the experience of mis- and disinformation, using both quantitative and qualitative methods.
 - To provide independent monitoring of the news media industry's capacity to sustain trustworthy news, focusing on threats to legacy and online news media, and how the platformization of news is affecting news media organizations' financial sustainability.
 - To study different types of mis- and disinformation, including hate speech and conspiracy theories, and their impact on the public sphere, focusing on actors in addition to the far right, and on how they are incentivized to weaponize information.

- To examine how different news media formats can be used to reach a broad public, and how these formats are perceived with a view towards reducing problems created by news avoidance and by selective news exposure.
- To provide holistic assessments of resilience to mis- and disinformation across all countries to inform strategies for countering mis- and disinformation and the conditions that give rise to these types of information.

CHAPTER 3

ARTIFICIAL INTELLIGENCE, INFORMATION ECOSYSTEMS AND DEMOCRACY





This map represents a statistical summary of the thematic content of this chapter. The network graph represents relations between the words in the chapter, placing them closer to each other the more they are related. The bigger the node, the more present the word is, signalling its role in defining what the report is about. The colors represent words that are closely related to each other and can be interpreted as a topic.

The map is generated by the OID on the basis of the chapter's text using GarganText – developed by the CNRS Institute of Complex Systems. Starting from a co-occurrence matrix generated from chapter's text, GarganText forms a network where words are connected if they are likely to occur together. Clustering is conducted based on the Louvain community detection method, and the visualization is generated using the Force Atlas 2 algorithm.

[Link to the interactive map here](#)

This chapter examines research on the properties of AI systems (specifically machine learning algorithms) and how they are embedded in online content governance systems. It is essential to understand these systems if violations of human rights are to be reduced and flows of mis- and disinformation are not to become an even greater threat to information integrity and to the health of information ecosystems.

The research synthesis focuses on:

- **How is ‘artificial intelligence’ (AI) defined, and what are the relationships between AI systems development and internationally protected human rights?** The chapter explores whether new rights are needed as AI systems become widely used, and examines the challenges presented by biases in the inputs and outputs of large language models (LLMs). The implications of AI systems for fundamental rights, including freedom of expression and information, privacy and democratic participation, are addressed.
- **What impact do AI systems and content governance, including content generation and content moderation and curation, have on information integrity?** Attention is given to the technologies used for content governance. The use of generative AI (GenAI) by mis- and disinformation actors is also discussed, together with assessments of approaches to countering this type of information and the impacts of generative AI and algorithmic content curation systems on the news media industry.
- **What are the interdependencies between AI systems development, the use of automated tools and democratic processes?** The consequences are discussed, including the influence on debate in the public sphere, the impacts on societal resilience and social sustainability and on environmental sustainability.

The chapter provides a comprehensive assessment of research in these areas, highlighting both the benefits and risks to the health of information ecosystems.

Further discussion of AI systems occurs in later chapters. In Chapters 6 and 7, approaches to AI systems governance that are being put into place by governments, tech companies and not-for-profit organizations are examined. Chapter 8 turns to why the increasing dependency on AI systems and data extraction and processing produces discriminatory outcomes and to strategies aimed at reimagining and practicing alternative approaches to data governance.¹

¹ For background on AI systems governance, see Bullock *et al.* (2022); Gunkel (2024); Paul *et al.* (2024); Quintavalla & Temperman (2023). For a review of advances in research on generative artificial intelligence (GenAI), including challenges and threats, see Bontcheva *et al.* (2024). See Appendix: Methodology for details of literature review process.

1 Introduction

Humans are social beings. They communicate to achieve common goals, based on convictions they develop through information they receive and share. Democratic decision-making processes cannot function without public discussion of questions of general interest, sharing of ideas and debate about proposed courses of action and past decisions. These processes have become heavily digitalized (i.e., taking place in online spaces) and mediatized (taking place in, and under, the rules, practices and algorithmic systems of privately owned communication spaces). These spaces – information ecosystems – have rules, just as offline spaces do. In offline public spaces, laws set by states and enforced by executive power define the rules for public debate. In online settings, the rules under which communication takes place are set primarily by private actors, such as the owners of the digital platforms in which they take place, within the limits of what the laws allow. These actors enforce their communication rules through systems for content moderation that determine if the content is in keeping with the rules, and curation that decides how to direct the content to platform users. The more platforms seek to automate these systems through the use of artificial intelligence (AI),² the more they impact online communication processes and, ultimately, influence democratic discourses and democracy. The integrity of information ecosystems therefore depends on an environment that favors transparency and accountability.³

No single definition of ‘AI’ is accepted by all.⁴ The European Union’s Artificial Intelligence Act of 2024, for example, defines AI systems as:

A machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.⁵

The OECD definition is similar.⁶ Neither of these definitions claims that AI systems emulate human intelligence. Instead, the focus is on functional capabilities that derive from using machine learning (ML) algorithms that work by identifying patterns in data. This interpretation is reinforced by a report prepared for a European Commission Joint Research Centre:

AI is a generic term that refers to any machine or algorithm that is capable of observing its environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions. There are many different technologies that fall under this broad AI definition. At the moment, ML [machine learning] techniques are the most widely used.⁷

This definition is interesting because it makes explicit the technologies – that is, algorithms, ML – that constitute AI systems and that other definitions gloss over. It is a reminder of the need to ‘look under the hood’, to challenge ‘the thingness of AI and its status as a stable and agential entity... To let the term pass is to miss the opportunity to trace its sources of power and to demystify its referents’.⁸ It is therefore important to engage in a critical discussion on what ‘AI’ is. However, two factors

² *The Eye of the Master* presents a social history of AI systems, emphasizing that they are not ‘intelligent’ and that work in this field has been motivated historically by interests in labor saving and surveillance (Pasquinelli, 2023). There are many warnings about the inherent problems in anthropomorphizing AI systems (Floridi & Nobre, 2024). There are suggestions for a new glossary of terms, for example, ‘systems for statistical propositions’, to describe large language models (LLMs) to support discussion of the benefits and harms of technological advances more transparently (Frau-Meigs, 2024b). In the field of political communication, for example, ‘AI’ has been defined as ‘the tangible real-world capability of non-human machines or artificial entities to perform, task solve, communicate, interact, and act logically as it occurs with biological humans’ (Gil de Zúñiga et al., 2023, p. 2), supported by the Spanish National Research Council (CSIC, Consejo Superior de Investigaciones Científicas). Activists and critical scholars emphasize the importance of focusing not just on technology, but also on politics, power structures, cultural narratives and public perceptions (PublicSpaces International, 2024; Verdegem, 2021).

³ Nowotny (2021); Puddephatt (2021).

⁴ Samoli et al. (2020).

⁵ EC (2024c, Article 3(1)).

⁶ An AI system is ‘a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment’ (OECD, 2022c, p. 7).

⁷ Annoni et al. (2018, p. 18).

⁸ Suchman (2023, p. 1).

make the use of the term 'AI' in this report difficult to avoid; first, the proprietary nature of many systems means that details of the technologies used are often not disclosed; and second, 'AI' is widely used, not only in the research literature, but also in both popular and official discourses.

So-called 'generative AI' (GenAI) refers to a broad category of ML systems that are capable of synthesizing content. They are typically trained on very large data sets and can generate content – synthetic media – in the form of text, images and video that may often be difficult to distinguish in terms of quality from human-generated content. Among the various examples of GenAI systems, large language models (LLMs) are the best known. Despite being classified as GenAI, however, LLMs are simply statistical models of language use. While systems that use LLMs, such as chatbots, can produce very plausible responses to queries, this should not be mistaken for *natural language understanding*. An LLM, then, is: 'a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a *stochastic parrot*'.⁹

LLMs first achieved public attention in November 2022 with the announcement of ChatGPT by OpenAI, and are already being used in ways that have significant implications for the public's experiences of information ecosystems and the content that diffuses through them. These include, for example, to create and moderate content such as hate speech; to create realistic 'deepfakes', but also to detect them; and to promote, but also to fight, mis- and disinformation.¹⁰ And as the realism of 'deepfakes' increases, their detection becomes correspondingly harder.¹¹

In its 2024 report, *AI as a Public Good: Ensuring Democratic Control of AI in the Information Space*, the Forum on Information and Democracy said that AI systems, particularly GenAI systems, are 'revolutionizing the way we create information across various mediums, including text, audio, images and video, presenting both challenges and opportunities'.¹² Gaining democratic control of AI systems requires effective accountability structures for the whole AI systems lifecycle, which the OECD defines as setting objectives and the functional specification, building a model to meet the specification, and its verification and validation as well as its deployment, operation and monitoring.¹³

It will be clear that there is not *an* AI; rather there are different ML technologies, instances of which may be involved in processes related to information retrieval, synthesis, presentation and governance. ML technologies vary widely, ranging from relatively simple algorithms executing tasks (such as filtering for specific words) to deep learning algorithms (that can be trained to assess the likelihood of content having been authored by an inauthentic actor, such as a disinformation bot).¹⁴

Embedding AI systems within information ecosystems impacts on content production (synthesizing text, images and video), moderation (deciding if content violates regulations) and consumption (deciding on the content's audience). It is therefore unsurprising that concerns have been raised about the potential for these systems to shape public discourse and, moreover, to do so in ways that may have significant implications for societal cohesion and resilience.¹⁵ Questions about the use of 'AI' in information ecosystems cannot be settled on technical criteria alone, but must address a much broader range of issues, including legal (e.g., does their use discriminate against certain groups?) and societal (e.g., does their use reduce the diversity of information available to publics?).¹⁶

⁹ Bender et al. (2021, p. 617), supported by the National Science Foundation (NSF), US.

¹⁰ See Bonfanti (2020), Real Instituto Elcano, independent think tank; Kertysova (2018); Spitale et al. (2023).

¹¹ Ghosal et al. (2023). The prevalence of 'deepfakes' and other types of mis- and disinformation is discussed in Chapter 5, and efforts to combat them are discussed in Section 3.5 of this Chapter, and more extensively in Chapter 7.

¹² Forum on Information and Democracy (2024a, p. 18).

¹³ OECD (2023).

¹⁴ Veale et al. (2023).

¹⁵ De Gregorio & Stremlau (2023), supported in part by the European Commission.

As the capabilities of AI systems continue to advance and find application within information ecosystems, it is anticipated ‘that algorithmic moderation and regulation will become more and more seamlessly integrated into our social lives’.¹⁷ As this process progresses, the increased ‘consumption and commodification of artificial intelligence applications in daily life’, coupled with the ‘extensive trust and reliance on these technologies in public and private sectors’, makes it essential to confront important rule of law questions.¹⁸ The many different ways in which these and other questions may be answered should act as a timely reminder that how new technologies become embedded within people’s everyday lives is neither inevitable nor identical in different countries and regions, and can be shaped and influenced through normative choices based on ethical values and societal (and international) goals to be pursued (or not), as well as the experience and outlook of people in different regions and countries.

Discussions about how to ensure the health of information ecosystems that increasingly depend on AI for their day-to-day function need to be as inclusive as possible. While the Global North deals with the effects of the fast-growing pace of technological change on information ecosystems, the Global Majority World struggles with issues such as access to the internet, inequalities in investment in online safety and content governance resources, poor infrastructure and weak technology literacy levels.¹⁹ This means that some parts of the world are excluded from experiencing the benefits of AI systems (as well as other components of the digital infrastructure). As Kenichiro Natsume, Assistant Director-General at the World Intellectual Property Organization (WIPO), pointed out, ‘[the] 2.6 billion [unconnected] people [who] are not part of the digital world ... are [also] not part of the AI world’.²⁰

Exclusion from the internet keeps this population, which is disproportionately located in the Global Majority World, from accessing online information, and also from using AI tools, including GenAI.²¹ This does not mean these populations are unaffected by ‘AI divides’ since they are recipients of information that circulates by other means. Even when internet connectivity is achieved and affordable, the terms and conditions of online information access are skewed and shaped by big tech companies and by communication infrastructure providers that influence what information can be accessed, and which information is amplified by AI systems use and algorithms for those who are connected.²²

2 AI Systems and Human Rights

This section examines how human rights apply in the digital age, the problems created for fairness as a result of algorithmic biases, the importance of freedom of expression and information as well as privacy protection in considering the impacts of AI systems developments, and the impact on participatory rights as a result of the use of AI systems to manipulate information.

2.1 NEW TECHNOLOGIES – BUT NO NEW RIGHTS

Human dignity serves as the cornerstone of human rights. Thirty years ago, the guiding principles of the Vienna Declaration on human rights emphasized the indivisibility, universality, interrelatedness, and mutually dependent and reinforcing nature of all human rights.²³ Predating this, the *Universal Declaration of Human Rights* (UDHR) of 1948

¹⁶ Katzenbach (2021). This is especially so when companies such as OpenAI put ‘shiny products’ above safety, as claimed by researchers who have since left the company (Milmo, 2024).

¹⁷ Katzenbach (2021, p. 6).

¹⁸ De Gregorio (2023, p. 1).

¹⁹ De Gregorio & Stremmlau (2023), supported in part by the European Commission.

²⁰ Quoted in Vanoli (2024).

²¹ Fendji (2024). Some have limited access by sharing internet access accounts, but others have no internet access, due to absent or costly infrastructure. See Heeks (2022), supported by the Economic and Social Research Council (ESRC), UK; Mutsaers & Ragnedda (2019); Okolo (2023).

²² This issue is discussed further in Section 4.1, Chapter 6.

²³ OHCHR (1993).

committed states to the ‘promotion of universal respect for and observance of human rights and fundamental freedoms’, declaring these rights a ‘common standard of achievement for all peoples and nations’.²⁴

Human rights are fully applicable in the age of digital transformations, although much work is needed to uphold them in practice. As the then-United Nations High Commissioner for Human Rights, Michelle Bachelet, concluded in a speech in 2019, technology change does not necessitate new human rights conventions, but rather: ‘adapting the way we use institutions and processes... We can protect rights effectively only if we constantly fine-tune our processes to find the right mix of interventions’.²⁵

All societal actors have human rights obligations, albeit to differing degrees. Private online communication platforms have duties under the so-called Ruggie Principles, the *Guiding Principles on Business and Human Rights: Implementing the United Nations ‘Protect, Respect and Remedy’ Framework*.²⁶ Private entities need to protect, respect and provide remedies for violations of human rights, under the overall control of states. Following international human rights law, states have to respect, protect and ensure these rights for anyone within their control or jurisdiction,²⁷ although in the absence of regulation, these duties are not necessarily binding on all actors.

While digital platforms tend to frame their mission in human rights terms, such as ‘giving people a voice’ or ‘protecting expression’, the focus of research has been primarily on potential human rights violations by governments and less on areas

where platform business models might negatively impact user rights.²⁸ In the light of a reluctance to commit to substantial transparency obligations over the last decades, regional and national approaches have emerged to apply human rights obligations more directly to platforms.²⁹ The first ‘big picture’ approach can be seen in the European Union’s Digital Strategy, which attempts to curtail the influence of large digital companies by imposing obligations on them that mitigate the negative effects of online communication and, at the same time, promote the implementation of fundamental rights.³⁰ The European Union has emerged as a key international norm-maker for the digital arena, sometimes referred to as the ‘Brussels Effect’.³¹ Legislation, including the AI Act of 2024, provides some substantive obligations, but through stringent transparency and compliance obligations.³² Selected human rights issues that arise in the context of automated content governance and that impact on democratic decision-making processes are outlined below.

United Nations initiatives, such as a March 2024 General Assembly Resolution, show how there is awareness of technology’s role in both contributing to disruptive change and having the potential to build bridges within and between countries. The Resolution emphasizes that trustworthy AI systems for sustainable development should be promoted globally in line with existing human rights obligations.³³ By September 2024, AI systems had been positioned with other frontier technologies as a means to ‘turbocharge development’, securing a place as Objective 5 of the United Nations’ Global Digital Compact, which emphasizes the need for a ‘balanced, inclusive and risk-based approach to the governance of artificial intelligence (AI)’.³⁴

²⁴ UN (1948, preamble).

²⁵ Bachelet (2019).

²⁶ Ruggie (2011).

²⁷ Fischer-Lescano (2016), funded by the European Research Council (ERC).

²⁸ Jørgensen (2017); Kettemann & Schulz (2023).

²⁹ Müller & Kettemann (2024).

³⁰ EC (2022b).

³¹ Bradford (2020).

³² EC (2024c); Müller & Kettemann (2024); Werthner *et al.* (2024). Governance arrangements for these technologies are discussed in Chapters 6 and 7.

³³ UN (2024c), adopted by the UN General Assembly on 21 March 2024.

³⁴ UN (2024b, pp. 41, 52).

2.2 ALGORITHMIC BIAS AND FAIRNESS

Algorithmic bias involves systematic errors within AI systems that lead to unfair results.³⁵ Unfairness can be understood as privileging, without adequate reasons, members of one group over another. When used in settings where automated decisions impact individual or collective rights or values, these biases can lead to unfair and untransparent outcomes, not least because of economic incentives to favor results consistent with corporate interests.³⁶ Algorithmic decision-making in areas such as employment, law enforcement and lending can disproportionately negatively affect marginalized communities, and contribute to their exclusion from participation in democratic processes or the full enjoyment of their rights.³⁷ Research predominantly in the Global North, but also in the Global Majority World, reveals how algorithmic bias can lead to decisions by law enforcement authorities that disproportionately penalize minority ethnic groups and immigrant communities.³⁸

Biases arise from various factors linked to how, by whom and in which institutional or organizational setting an AI system is developed, particularly regarding the data used to train it – for example, when training data is incomplete or contains historical prejudices or assumptions that are then replicated: if, in the text on which an LLM is trained doctors are primarily described as male, then answers generated by the LLM will replicate this.³⁹ Similar replication of stereotypes has been shown to happen in image-generating LLMs. Even AI systems trained using what is believed to be unbiased data may produce biased outputs, since a lack of transparency in how their outputs are produced may make it difficult to exercise effective oversight over their performance.⁴⁰ The personalization

algorithms used on social media platforms to decide what content users are exposed to exploit the data users create when they interact with content. Once ‘datafied’⁴¹ in this way, AI algorithms can be used to model user behavior, and the model can then be applied in ways that are biased towards the interests of platforms, leading to the promotion of content that maximizes user engagement at the expense of quality and veracity.⁴² These are all consequences of the way that LLMs synthesize their training data to produce outputs based on statistical prevalence – reducing the diversity of inputs into the specificity of a single output. In addition, LLMs may be trained on synthetic data, that is, ‘data that mimic and substitute empirical observations without directly corresponding to real-world phenomena’.⁴³ Critical assessments of the use of such data may be helpful in protecting privacy and improving data sets that have a representational link to the ‘real-world’, for example, addressing biases, but when developed by artificial neural networks this does not provide a means of explaining why a given output has been generated.

Algorithmic fairness refers to the aspiration of creating and implementing AI systems that do not discriminate or bias against specific persons or groups based on protected characteristics, such as race, gender or ethnicity.⁴⁴ Fair AI algorithms would make decisions without favoring one individual or group over another.⁴⁵ To achieve this, attempts are now being made to increase the quality of training data sets. IBM launched a Diversity in Faces data set to help overcome specific biases in facial recognition technology.⁴⁶ This data set includes a million images of faces annotated with details that provide a broad representation of human faces, such as age, gender, skin tone and facial

³⁵ Hasimi & Poniszewska-Marañda (2024); see further discussion of fairness Sections 2 & 3, Chapter 4 and in Chapter 8.

³⁶ The biases of personalization systems and search engines have been recognized in the literature and demonstrated empirically for at least a decade (Eubanks, 2018; Rieder & Sire, 2014).

³⁷ Baecker *et al.* (2023).

³⁸ Chouliraki & Georgiou (2022); Gurumurthy & Chami (2019).

³⁹ Belenguer (2022).

⁴⁰ Pollicino & De Gregorio (2022).

⁴¹ ‘Datafied’ means turning a previously computationally invisible activity into data, and is a term used especially in the literature that is critical of the datafication of the lives of human beings (van Dijck, 2014).

⁴² Pfeiffer *et al.* (2023), funded by Projekt DEAL, Alliance of Science Organizations, Germany.

⁴³ Offenhuber (2024, p. 1), and for a discussion of a variety of types of synthetic data and their implications.

⁴⁴ Ferrara (2024a); Johnson (2023).

⁴⁵ Hall & Ellis (2023).

⁴⁶ Smith (2019); the author was an IBM employee.

features drawn from many different countries and cultures. By using this data set, developers can train facial recognition systems that are less likely to reproduce stereotypes regarding certain groups.⁴⁷ This approach is based, however, on the premise that greater diversity will reduce the prevalence of bias, but can be limited by the unavailability of more diverse training data. The challenges around guaranteeing fairness will increase as AI progressively becomes enmeshed in the processes that define the social conditions in which meaning is produced. These, in turn, are dependent on the level of trust in them, their prevalence and institutional roles.⁴⁸

Diversity in training data is expected to contribute to mitigating the risks of bias in AI systems that use these models. Diversity in development teams can offer a variety of perspectives that challenge conventional norms and biases that may be overlooked in more homogenous teams. The setup of development teams – and those working on AI ethics generally – is substantially linked to product design decisions.⁴⁹ Microsoft has embraced this strategy through its Inclusive Design Initiative, which employs people with diverse backgrounds (including disabilities) to design and test new products.⁵⁰ Evidence of the effectiveness of such corporate diversity strategies is inconclusive, and in some cases no direct association is found between the socio-demographic diversity of AI systems developers and AI systems output biases. The viewpoint diversity of those holding ML, coding or data analyst jobs is found to play a much stronger role based on a relatively small-scale study.⁵¹ Various forms of discrimination are likely to persist in the prevailing culture, which is likely to be encouraged if its leadership is skewed to favor certain groups, as illustrated by high-profile resignations from some of the big tech companies.

In many countries of the Global Majority World, AI systems development and deployment are at a 'nascent stage', potentially allowing countries to design robust anti-discrimination rules before broad uptake.⁵² For the Global Majority World, questions about 'human rights, democracy and autonomy in the countries of the majority world are not trivial'.⁵³ For example, the development of fair AI systems may be hindered by the limited availability of training data in many Global Majority World languages.⁵⁴

2.3 FREEDOM OF EXPRESSION AND INFORMATION

Freedom of expression is a 'cornerstone' for the formation of democratic societies, and as such is protected by all human rights instruments, including Article 19 of the UDHR and Article 19 of the *International Covenant on Civil and Political Rights* (ICCPR), and all regional human rights conventions.⁵⁵ This right includes the freedom to express and hold one's own opinions, to impart information, to seek and receive information and, implicitly, freedom of media expression. Given the technological realities of online communication, the right to freedom of expression is implicated in other rights such as the right to health (seeking and imparting health-related information) and to education (seeking and imparting information related to education, attending classes and research papers).

AI systems allow for much easier access to online communication spaces and information interchange, but also impact what information can be seen.⁵⁶ All platforms use AI systems to govern online communication and optimize user engagement.⁵⁷ There is thus a substantial impact, across information ecosystems, of these content governance systems on freedom of expression.⁵⁸

⁴⁷ Wiggers (2019).

⁴⁸ Pfeiffer *et al.* (2023).

⁴⁹ Martin (2022).

⁵⁰ Microsoft (2023).

⁵¹ Chi *et al.* (2021); Harris (2023); Park (2024).

⁵² Gurumurthy & Chami (2019, p. 9).

⁵³ Ricaurte (2022, p. 732).

⁵⁴ Ricaurte (2022), citing Horowitz (2021); more recently, see HRW (2023).

⁵⁵ UN (1948, 1966).

⁵⁶ Dias Oliva (2020).

⁵⁷ Gillespie (2020); Longo *et al.* (2024).

⁵⁸ De Gregorio & Dunn (2023).

Figure 3.1
Illustration of user engagement



Source: Pixabay

Information enables individuals to make educated judgments by helping them become acquainted with facts (see Figure 3.1) and societal issues.⁵⁹ It is a crucial component of individual liberty. Nevertheless, people's ability to obtain and interpret information can be restricted when they encounter mis- or disinformation or biased content.⁶⁰ Another challenge people face is a lack of reliable, accurate information, a problem sometimes compounded by information overload, and worsened when there is a decline in global trust in news, which is associated with the prevalence of online mis- and disinformation, as discussed in Chapter 2.⁶¹

Focusing mainly on tweaking content governance practices and systems ignores the underlying causes of social discord and distrust that give rise to polarized public opinion. Some argue that a focus on the 'public worthiness' of information, rather than on information 'disorder', can reveal the complex elements of visibility, access, reflexivity, mediation, influence and information legitimacy. Better insight into how these can combine in different ways to

foster new imaginings of publicness could enable democracy to flourish.⁶²

2.4 PRIVACY PROTECTION

AI systems present significant challenges to people's right to data protection and privacy. The comprehensive collection and analysis of data by these systems leads to the development of multidata, points-based profiles of individuals, often without their explicit consent, which can separately and in aggregate violate their right to privacy.⁶³ For example, Meta has said it is extending the jurisdictions in which it collects public data to train its models beyond the United States, although Data Protection Authorities in the European Union and Brazil and beyond have sought to stop this practice. In other countries, such as the United Kingdom, after some changes this practice has been deemed a 'legitimate interest' in processing data.⁶⁴ Moreover, the way consent is obtained for data collection does not often meet the threshold of being 'informed'. Many users experience consent fatigue, agreeing to privacy policies without understanding the implications.⁶⁵

As most online communication takes place in private communication spaces that are financed through data collection, there is an incentive for platforms that use automated content governance tools to configure them in a way that maximizes data collection. This can lead to interferences with, and violations of, rights to privacy and data protection. These can be addressed to some extent by enforcing existing privacy and data protection laws and international human rights standards that emphasize consent, data minimization and purpose limitation in data processing – for example, Article 12 of the UDHR and Article 17 of the ICCPR protect privacy and personal data.⁶⁶

⁵⁹ Masur (2020).

⁶⁰ Measurement issues around the scale of mis- and disinformation are discussed in Section 2, Chapter 5, along with issues of public awareness of its prevalence in Section 3, Chapter 5.

⁶¹ Samoilenko & Suvorova (2023). The big tech platform's practices of reducing or amplifying content, for example, of reducing news media or user-generated content, are discussed in Chapter 2 (Table 2.1), Chapters 6 and 7 as a self-regulatory strategy.

⁶² Splichal (2022a); Geiß *et al.* (2021, p. 683), supported in part by the Media Authority of North Rhine-Westphalia, Germany.

⁶³ Bontridder & Poullet (2021).

⁶⁴ Forum on Information and Democracy (2024b).

⁶⁵ Abdulrauf & Dube (2024); Barocas & Nissenbaum (2014); Richards & Hartzog (2019); Turow *et al.* (2023); Avle (2022), supported by the National Science Foundation (NSF), US.

⁶⁶ UN (1948, 1966).

Training an LLM on personal data effectively encodes aspects of this data into the model's parameters. Even where there is no direct retention of the data,⁶⁷ the model learns patterns and information during training that may be reconstructed or inferred by analyzing its output. Studies have shown that it is possible to extract specific data points from LLMs through techniques like model inversion or membership inference attacks, where queries to the model can reveal if certain data was used in training.⁶⁸ However, this approach may be limited by the fact that it suggests that it is possible to invert outputs to inputs, ignoring that the model combines inputs according to probabilistic weights that are derived from a combination of inputs, rather than linearly from any single input. In addition, models learn and change when model–user interaction or ‘user embedding’ occurs in addition to learning in response to user text prompts.⁶⁹

There are several potential technical solutions and strategic reforms that can be implemented to address the privacy risks posed by LLMs and other types of AI. These aim to enhance privacy protection, ensure transparency and uphold ethical standards within AI systems development and deployment.⁷⁰ Regarding technical solutions, protection from de-anonymization risks can be achieved by using differential privacy methods that add random noise to the data in a way that prevents the identification of any individual from the data set, and these methods have been adopted by companies such as Apple and Google.⁷¹

Establishing and adhering to ethical standards when developing AI systems is essential to mitigate risks related to privacy, bias and other potential harms. For instance, the Institute of Electrical and

Electronics Engineers (IEEE) has proposed an ethical framework for AI and autonomous systems that includes guidelines for prioritizing human well-being, data agency and accountability in AI systems.⁷² Similarly, the Partnership on Artificial Intelligence to Benefit People and Society (Partnership on AI), which includes stakeholders from various organizations, promotes best practice in AI development, focusing on fairness, transparency and accountability in an effort to ensure AI systems are used responsibly, although substantial changes in company policies or product priorities have not materialized.⁷³

Legislative tools such as the European Union's General Data Protection Regulation (GDPR), its AI Act or the United States' California Consumer Privacy Act (CCPA) provide foundational frameworks for regulating AI systems and data use practices, aiming for comprehensive protection for individuals' privacy.⁷⁴ This approach includes strict requirements for transparency and data quality, setting a benchmark for global AI regulations, aiming to empower consumers with more control over the personal information that businesses collect about them – including transparency about data use and the right to delete collected personal data. Research suggests, however, that both the GDPR and CCPA have significant limitations and, in the case of the GDPR, regarding informing data subjects about how their data is being used.⁷⁵

Surveillance is defined as the ‘process of observing individuals or groups for a purpose and make inferences/judgements on their behavior’.⁷⁶ Its scope and scale have been transformed by ‘datafication’, that is, the quantification of people's everyday activities in real time by digital platforms.⁷⁷ AI systems algorithms can then be used to analyze this data to identify patterns of behavior. The risks

⁶⁷ This will depend on whether training requires access to personal data and where this data is stored. In the case of ChatGPT, any additional data must be uploaded to OpenAI's servers and OpenAI retains this data. Some LLMs allow for data to be retained locally.

⁶⁸ Jagannatha *et al.* (2021).

⁶⁹ Ning *et al.* (2024).

⁷⁰ Lepri *et al.* (2018); Yan *et al.* (2024); Ong *et al.* (2024), supported in part by the Wellcome Trust.

⁷¹ Zhao & Chen (2022).

⁷² IEEE (2019); Gunkel (2024); see also UNESCO's recommendation on the ethics of AI (2022c).

⁷³ Borocas *et al.* (2023); Caton & Haas (2020).

⁷⁴ EC (2016b, 2024c); Mahler (2022); US State of California (2018). See also Section 4.2, Chapter 6.

⁷⁵ Lee (2024); Wulf & Seizov (2022), supported by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) and Dutch Research Council (NWO, Nederlandse Organisatie voor Wetenschappelijk Onderzoek).

⁷⁶ Fontes *et al.* (2022, p. 2).

⁷⁷ ‘Datafication’ refers to the ‘transformation of social action into quantified data’ for real-time tracking and prediction (van Dijck, 2014, p. 198).

of – and potential remedies to – surveillance in the data economy, sometimes known as ‘dataveillance’, are discussed further in later chapters.⁷⁸

2.5 DEMOCRACY AND PARTICIPATORY RIGHTS

While United Nations member states are committed to democratic governance, the majority of the world’s population live in states that suffer from democratic deficits. A 2023 analysis of the state of democracy globally concluded that it is ‘complex, fluid and unequal’.⁷⁹ The *Vienna Declaration and Programme of Action* of 1993 clearly links democracy with human rights, urging member states to continuously foster democratic principles to enhance human rights protection.⁸⁰ Democracy fundamentally relies on the principles of free, equal, secret and independent elections and democratic decision-making processes. Particularly problematic are targeted mis- and disinformation campaigns that aim to manipulate elections and stir conflict.⁸¹ The manipulation of democratic decision-making processes is facilitated by AI systems. Even non-subliminal techniques can be manipulative, especially for vulnerable groups such as children, if they exploit mental health vulnerabilities, immaturity or lack of digital literacy. As AI systems evolve, the potential for misuse increases. A notable example is OpenAI’s Sora, one of several GenAI systems capable of producing video footage from minimal text input.⁸²

The utilization of AI systems in political campaigns and election processes leads to concerns about the transparency, accountability and manipulation of democratic decision-making. The electoral landscape faces significant risks from the very tools that enable campaigns to target voters with exceptional accuracy: the capacity to disseminate false information, to manipulate perceptions through microtargeting and to magnify

controversial content.⁸³ The capacity of AI chatbots, for example Microsoft’s Bing Chat, was tested over several months during elections in Germany and Switzerland. This GenAI chatbot produced factual errors to queries on election topics with a near 30% error rate.⁸⁴ The use of AI systems to personalize content on social media platforms has the potential to sway voters and create divisions in public opinion, affecting individuals’ abilities to freely engage in their government and public affairs. Advanced data analytic capabilities have made voter microtargeting significantly more accessible. While this has the capacity to enhance engagement and voting percentages, it also exposes voters to manipulation via hyper-targeted content that can seek to sway their opinions or even discourage them from voting.⁸⁵

3 AI Systems and Content Governance

AI systems deployed by digital platforms manage the visibility and spread of information, mis- and disinformation.⁸⁶ This section addresses content generation and governance, that is, content moderation, distribution and amplification; it assesses the impact of AI systems on information ecosystems; and discusses how AI systems are being used by mis- and disinformation actors.

Social media platforms have become vital arenas for public debate, where users gather information, share ideas and form opinions. Content governance systems impact on these processes because they frame the conditions under which content is seen and with whom it is shared.⁸⁷ These systems utilize

⁷⁸ ‘Dataveillance’ refers to continuous surveillance using (meta)data (van Dijck, 2014). Surveillance is examined further in Chapters 4 and 8.

⁷⁹ International IDEA (2023).

⁸⁰ UN OHCHR (1993).

⁸¹ See Section 4.3.3, Chapter 2 for a discussion of the weaponization of information and election manipulation.

⁸² Liu *et al.* (2024); one of the authors works with Microsoft Research.

⁸³ Schippers (2020).

⁸⁴ See Helming (2023). The impact of mis- and disinformation on political processes is discussed in Section 3, Chapter 2 as well as in Chapter 5, where AI literacy and capacities to discern accurate from inaccurate information, including the ‘hallucinations’ generated by AI systems, are discussed.

⁸⁵ Michael (2023); the Cambridge Analytica story is discussed in Section 4.3.3, Chapter 2.

⁸⁶ Sančanin & Penjišević (2022).

user behavior, previous choices (interest histories) and past interactions to customize content streams, control content visibility and enhance engagement metrics. AI-based content governance systems are intended to reduce the prevalence of undesired content such as mis- or disinformation, including hate speech and propaganda.⁸⁸ Importantly, their design, implementation and accountability lie in the hands of the platforms where they are used; these systems, and the governance policies and practices they are intended to support, vary from platform to platform and from jurisdiction to jurisdiction, changing over time, especially with changes in ownership, as illustrated in the case of X/Twitter.⁸⁹

Twitter's transformation under Musk. On his takeover of Twitter in late 2022, Elon Musk announced: 'The reason I acquired Twitter is because it is important to the future of civilization to have a common digital town square, where a wide range of beliefs can be debated in a healthy manner, without resorting to violence'.⁹⁰ Not long after, he introduced significant changes to Twitter's content policies and practices, signaled by reinstating some high-profile users who had been banned for violating the platform's misinformation and hateful conduct policies.⁹¹ Changes in practices were inevitable, with the sacking of a large proportion of staff responsible for human rights, AI ethics, trust and safety.⁹² X introduced Community Notes, which aim 'to create a better informed world by empowering people on X to collaboratively add context to potentially misleading posts',⁹³ but retains control over which contributions are approved and made visible to users.

The Australian eSafety Commissioner has criticized X for letting the worst offenders back online, 'while at the same time significantly reducing trust and safety personnel whose job it is to protect users from harm'.⁹⁴

The AI algorithms that drive social media platforms are designed to enhance user engagement by personalizing online experiences, and are, in principle, neutral on the veracity of content. However, if mis- or disinformation content provides the most engagement, the system – if not properly reviewed – will increase the dissemination of such content.⁹⁵

3.1 AI SYSTEMS IN CONTENT GENERATION

The availability and ease of use of GenAI has arguably 'democratized' content production. Making a video used to be the reserve of a privileged few. Without specific detailed technical know-how, users can now create digital content in audio, video or text, using a wide range of apps, and distribute them through digital platforms.⁹⁶ With this comes potential 'side effects', which stem from the increase in volume, velocity and potential persuasiveness of problematic content and its decreasing cost of production.⁹⁷

Digital platforms have started to address the challenges of text and speech produced by GenAI, but, in jurisdictions without rules on risk assessment obligations, the internal rules are often vague or inconsistently enforced: 'The driving force is either the misleading and harmful potential or a more compliance-oriented approach in terms of copyright and quality standards of the content'.⁹⁸

⁸⁷ Jungherr & Schroeder (2023), funded by the Volkswagen Foundation (Volkswagen Stiftung).

⁸⁸ Christodoulou & Iordanou (2021), funded by the European Commission.

⁸⁹ Burkart & Huber (2021); see EC (2024d), for demanding that X explain its content moderation compliance with European Union regulations.

⁹⁰ York (2022).

⁹¹ Ivanova (2022).

⁹² Brewster (2024); eSafety Commissioner (2024a).

⁹³ X (2024).

⁹⁴ eSafety Commissioner (2024b).

⁹⁵ Bontridder & Poulet (2021); Ohme *et al.* (2024); Reisach (2021); see also Chapter 2 where audience/user engagement with content and mis- and disinformation research is discussed.

⁹⁶ Allen & Weyl (2024); Cooke (2023).

⁹⁷ Feuerriegel *et al.* (2023).

⁹⁸ Miguel & Krack (2023, p. 3).

Given that audiences find it difficult to distinguish between GenAI and human-produced content,⁹⁹ it is important to raise levels of AI literacy and to impose disclosure obligations for AI-generated content or AI-operated accounts.¹⁰⁰ So far, however, despite high-profile cases, there is no evidence that GenAI is *systematically* used as a tool to synthesize politically motivated mis- and disinformation.¹⁰¹ Even if this is the case, there is no doubt that it is being used with growing pressures on tech companies, prompting them to sign a voluntary accord in early 2024 to prevent AI systems from disrupting elections.¹⁰²

Copyright is a challenging issue for AI-generated content because it applies both to the data used for training and to the generated output. Training AI systems, especially LLMs, often involves ingesting vast amounts of text harvested from the internet, much of which is copyrighted (although some exceptions exist), raising questions about whether this usage constitutes ‘fair use’ or ‘exceptions’ to copyright depending on the jurisdiction, or requires explicit permission from rights holders. The output generated by AI systems also sometimes (and inexplicably) reproduces copyrighted material verbatim.¹⁰³

3.2 AI SYSTEMS IN CONTENT MODERATION AND CURATION

AI systems are increasingly used by platforms for implementing content governance guidelines on how content is sourced (or created) and then distributed. Content moderation involves identifying and removing or flagging inappropriate, harmful or illegal content based on predefined criteria. This definition of criteria, the setting up of internal standards and community guidelines is a powerful

act, which, coupled with algorithmic content moderation and curation (i.e., governance), gives digital platforms a role that researchers call the ‘arbiters of truth’.¹⁰⁴ It is not so much ‘truth’ that is decided on, however, but what content stays on a platform and what content is given more visibility. Content curation systems then select and organize content that has passed the moderation stage for distribution. These systems are used by platforms to determine who sees what content, often personalizing it by matching against users’ preferences, as revealed by their past behaviors.¹⁰⁵ The use of these systems takes place within the framework of existing and new laws shaping platform behavior, including rules for transparency and user rights. No moderation or content curation system is neutral or non-discriminatory. If it did not treat content differently, it would not be doing its job. Certain categories and procedures must be used to structure the content presented to social media users. Choices must be made even if the choice is to present content in chronological order. As a report for UNESCO’s regional office in Montevideo put it:

AI technologies are not neutral; they inherently reflect the values of their developers and the broader development and deployment ecosystem. While they have the potential to enhance accountability in public institutions and their representatives, foster greater participation and pluralism to enrich citizen engagement, and make democracy more inclusive and responsive, they can also amplify autocratic tendencies and be used for potentially malicious and manipulative purposes.¹⁰⁶

⁹⁹ Kreps *et al.* (2022).

¹⁰⁰ AI literacy is discussed in Chapter 5.

¹⁰¹ Kreps *et al.* (2022); Simon *et al.* (2023).

¹⁰² O’Brien & Swenson (2024). This accord is discussed in the context of the governance of political processes in Chapter 7.

¹⁰³ Geiger (2024) discusses a human rights-friendly copyright framework for GenAI, emphasizing the rights of human creators. UNESCO began considering the impact of AI systems on cultural production earlier than the current debate about LLMs (Kulesz, 2018). WIPO states that there is significant legal uncertainty, and answers are likely to vary by jurisdiction (2024). In the European Union, if a work is created by AI, it is not subject to copyright, but there is scope for application of the law if a creator is deemed to have given explicit instructions to an AI application. The AI Act says that text and data-mining operations must receive consent unless they are subject to exemptions – which so far seem to apply – but companies must document their use of data and court proceedings are underway. As of August 2024, the United States does not offer copyright protection to creations produced by GenAI, and it is not clear what liability OpenAI and other firms have for scraping data to train LLMs. Legislation is being presented to Congress, but none has succeeded in becoming law. The issues in this area relating to ‘fair use’ in the United States, copyright exceptions in the European Union and provisions regarding text and data mining in the European Union, as well as whether news media organizations should be compensated for platform use of ‘snippets’ and other texts, are not examined in-depth in this report, but see Section 2, Chapter 2 and Section 4.5, Chapter 6 for a discussion on compensation.

¹⁰⁴ Schaafe & Fukuyama (2023); see Gillespie *et al.* (2023, p. 4), for an expanded research agenda on content moderation, arguing for grasping ‘the breadth and depth of moderation, across the entire ecosystem of content provision and deep into the infrastructural stack of distribution’.

¹⁰⁵ Gillespie (2020).

¹⁰⁶ Innerarity (2024, p. 10).

It is essential to puncture the “fallacy of AI neutrality” – represented by the mistaken belief that AI systems can be designed in an inherently unbiased and neutral manner¹⁰⁷. Research shows that content moderation and curation systems suffer from biases and encode non-transparent decision-making processes. They are optimized for engagement, that is, to personalize and distribute content to audiences that the system predicts will engage them meaningfully.¹⁰⁸ This means that they can be designed and deployed to achieve political ends, and in ways that exacerbate individual and societal risks. GenAI content is distributed in this way, even though a substantial number of platforms do not have sufficiently detailed policies in place, and may not adhere to them when they do.¹⁰⁹

There are concerns about the political sensitivity of LLMs and their potential to deepen societal divisions.¹¹⁰ These tools are relatively new and are being updated quickly, but research shows that the output of three LLM-based chatbots (ChatGPT, Bing Chat and Bard) seems to exhibit varying degrees of bias in response to political queries concerning authoritarian regimes. This is influenced by the language of the prompt. Significant disparities have been found regarding chatbot answers, with Russian language queries resulting in evasive answers regarding content that can be viewed as critical of Russian authorities. Anecdotal evidence shows that this applies to similar queries in Mandarin Chinese on issues such as the persecution of Uyghurs.¹¹¹

Human moderators still play a role in rechecking certain automated decisions and, depending on the jurisdiction a social media company operates in, become active once a user requests that a

content-related decision is reviewed.¹¹² These ‘cognitive assemblages’ involved in content moderation have been described as a ‘cobbled space of pre-emptive calculation’.¹¹³ In all, the trend clearly goes towards more automated moderation, especially in areas where the law is regarded as being clear, such as terrorism content.¹¹⁴ Where the law is less clear, as in the case of mis- and disinformation, automated tools focus less on content and more on markers related to the distribution channel or the behavior of the account from which the content was launched.¹¹⁵ Researchers criticize that even ‘well-optimized’ moderation and curation systems can ‘exacerbate, rather than relieve, many existing problems with content policy’ because they increase opacity, complicate ‘issues of fairness and justice in large-scale sociotechnical systems and ... re-obscure the fundamentally political nature of speech decisions being executed at scale’.¹¹⁶ As discussed, since content produced by AI systems can exhibit and/or reinforce biases against historically marginalized and minority groups,¹¹⁷ safeguards need to be implemented to prevent these systems from intensifying existing societal inequalities, along with efforts made to use these systems to help elevate the representation of underrepresented groups in the content produced.¹¹⁸ Efforts to promote ethical standards and diversity in development teams are part of the solution but are not themselves sufficient.¹¹⁹

The ‘hyper-personalization’ of content curation systems attracts much criticism in the literature. Some researchers fear that they may lead especially vulnerable media consumers, such as children and young adults, into ‘rabbit holes’ of potentially harmful content, among many other harms.¹²⁰

¹⁰⁷ Verhulst (2023, p. 1).

¹⁰⁸ Sančanin & Penjišević (2022).

¹⁰⁹ Issues of the weaponization of information are discussed in Section 4.3.3, Chapter 2, and impacts of content moderation practices are discussed in Section 2.3, Chapter 7.

¹¹⁰ Biju & Gayathri (2023).

¹¹¹ Urman & Makhortykh (2024).

¹¹² The role and effectiveness of human oversight is discussed in Section 2.1, Chapter 7.

¹¹³ Crosset & Dupont (2022, p. 10), supported by the Fondation du Risque (Allianz, Axa, Groupama and Société Générale) in partnership with the Institut Mines-Télécom and Sciences Po.

¹¹⁴ Haas & Kettemann (2024); Macdonald *et al.* (2019).

¹¹⁵ Bontridder & Poulet (2021).

¹¹⁶ Gorwa *et al.* (2020, p. 1), supported by the Social Sciences and Humanities Research Council (SSHRC) of Canada.

¹¹⁷ Ross Arguedas & Simon (2023).

¹¹⁸ Forum on Information and Democracy (2024a).

¹¹⁹ The weaknesses of these efforts are discussed in Chapter 8.

¹²⁰ Amnesty International & AI Forensics (2023). These issues are discussed in Section 4, Chapter 5.

‘Harmful’ is a difficult criterion to use as a basis for assessing platform content policies. For example, there are few globally accepted examples of prohibited speech. Much ‘hate speech’, for instance, falls under the protection of free speech in rules in some jurisdictions such as the United States and, depending on the jurisdiction, there are different definitions of illegal speech.¹²¹ This is why it is sometimes argued that automated systems would work better if there was global consensus or a largely agreed on definition of what to find or filter, as in certain cases of terrorism and terrorism financing.¹²² Any such effort to forge consensus is likely to be disputed due to cultural and political differences and, even if achieved in the framework of international human rights obligations, may not be translated consistently into practice.

AI systems have also been used, for example, to improve crisis communication.¹²³ A study by the Organization of Security and Cooperation in Europe (OSCE) suggested that states should mandate platforms to undertake ‘crisis-sensitive human rights due diligence’, ‘crisis-sensitive human rights risks and impact assessments’ and emergency measures. Any platform action should ‘consider proportionality and reliability on AI tools and automated measures’.¹²⁴ Globally, crisis-sensitive human rights approaches by private actors have been urgently demanded.¹²⁵ Similar obligations are outlined in Europe’s new digital rules, such as the Digital Services Act of 2022, which contains obligations for platforms to conduct risk assessments as to the impacts of their rules and moderation practices on values, including societal cohesion, public health and democratic decision-making processes.¹²⁶

Given that platforms use AI systems for content governance, it is best practice (and legally required in certain jurisdictions, such as the European Union) that they should inform users. However, research shows that users tend to trust moderation decisions less when they know they are automated.¹²⁷ This showcases the complexity of achieving the responsible visibility of automated content governance, and user trust is also conditioned by education background and the sociopolitical setting.¹²⁸

One approach is to increase meaningful oversight, including external control over algorithmic systems.¹²⁹ This intervention into the private communication realm by platforms, governed by terms of service and algorithmic systems, can be legitimized by reference to the increasing impact of these norms and practices on public values that need to be integrated into the systems. Expert panels or selected user groups, sometimes referred to as platform councils or social media councils, have been suggested. Meta’s Oversight Board is one of the early efforts to make the governance of a commercial platform more inclusive of external input.¹³⁰ The impact on Meta itself tends to be judged as largely positive, if not very effective, and the Board has been described as overseeing ‘one of the largest speech systems in history’.¹³¹ However, the Board has not had substantial cross-industry influence, and has been unable to substantially change the speech governance priorities that Meta exhibits.¹³²

3.3 AI SYSTEMS AND NEWS MEDIA

Content created by GenAI can benefit news media diversity by contributing to the efficiency of content generation in specific contexts, and by

¹²¹ Gillespie (2020); see also Galli *et al.* (2023).

¹²² Haas & Kettemann (2024).

¹²³ On the substantial field of research on crisis communication, including communication strategies using social media, see Coombs & Holladay (2022); Jin & Austin (2022). The impacts of social media on conflict escalation are discussed in Section 4.3.3, Chapter 2, on the weaponization of information.

¹²⁴ Haas & Kettemann (2024, p. 9).

¹²⁵ Fatafta (2024).

¹²⁶ This legislation is discussed in Chapters 6 and 7.

¹²⁷ Ozanne *et al.* (2022), funded by the US Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) project. See also Chapters 5 and 7.

¹²⁸ Kim & Moon (2021).

¹²⁹ Nahmias & Perel (2021). Impacts on polarization are discussed in Section 4.4, Chapter 2. Various forms of oversight, including fact-checking, are discussed in Chapter 7.

¹³⁰ Kettemann & Schulz (2023).

¹³¹ Douek (2024, p. 373).

¹³² Ang & Haristya (2024); Douek (2024); Gulati (2023). Boards such as Meta’s examine specific cases of content moderation judgments. Broader forms of oversight aimed at increasing accountability are limited by researcher access to relevant data is discussed in Section 3.5, Chapter 9.

supporting journalists in optimizing the circulation of content contributions after publication, since articles can be published in cross-media formats without substantial additional costs. Research points to the potential of GenAI to ‘synthesize broadcast videos using news text during a news broadcast’ with better results than manual generation.¹³³ However, the implementation of these systems requires time and investment, and gains in efficiency and productivity should not be assumed.¹³⁴ As the use of GenAI becomes widespread, this can alleviate the burden of relying on overworked newsrooms by automating certain, more mundane, reporting tasks. However, the challenges of news organizations’ use of these systems need to be addressed if trust in news media output is to increase, and these organizations are to adhere to ethical standards of data collection and the principle of universality, in contrast to promoting personalized news and other content.¹³⁵ AI systems also have a bearing on freedom of expression when they influence editorial decisions, especially when there is a conflict between the editorial need for autonomy and goals that AI tools are optimized for.¹³⁶

The adoption of AI tools by news media organizations for content creation is a concern due to the growing dependency of news media organizations on these technologies.¹³⁷

A survey published in 2023 indicates that GenAI tools, such as ChatGPT, were being used in 49% of newsrooms worldwide.¹³⁸ Countries in the Global North and China are leading innovation in AI newsrooms, and research on adoption mainly focuses on the Global North.¹³⁹

Once content is created, news organizations are increasingly dependent on the AI systems used by digital platforms for distribution or circulation. This dependence raises the need for attracting audience traffic that is stimulated by algorithmic personalization. The effects of the interaction between audience traffic and the means to increase the flow of this traffic have implications for the production and visibility of content.

The question of who or what curates content online takes some of the power away from the hands of journalists, the traditional gatekeepers. Platform selection mechanisms usually involve a combination of algorithmic curation (based on criteria specified by business managers) and human editors, making it unclear what the core values underlying selection decisions are, and to what extent they reflect core democratic principles.¹⁴⁰ The impact of curation systems is especially sensitive in public service media (PSM) environments that have a mandate to reach a broad public.¹⁴¹ Research conducted in France, Germany, Greece, Italy, Poland and Sweden emphasizes that the news media’s growing dependence on algorithms means that those who access news media online to meet their information needs do so despite their concerns about the risk of encountering mis- and disinformation.¹⁴² Certain platforms have started to deprioritize news and favor more personal or emotionalizing content.¹⁴³ Weaker distribution of accurate information is associated in some studies with more polarized and polarizing media consumption behavior.¹⁴⁴ Platform algorithms using AI tools play a big role in shaping news distribution.¹⁴⁵ It is clear that some news organizations depend heavily on online traffic driven by third-party digital services, leading to dependency on social media for news distribution,

¹³³ Wu *et al.* (2023).

¹³⁴ Simon (2024); Simon & Isaza-Ibarra (2023).

¹³⁵ Horowitz *et al.* (2022); Ross Arguedas & Simon (2023); Vaccari & Chadwick (2020). Issues of changes in journalism practices are discussed in Section 4.1, Chapter 2 and of news media content moderation in Section 3.2, Chapter 7.

¹³⁶ Helberger *et al.* (2020).

¹³⁷ Simon (2022); see also the survey of AI guidelines for media across 17 countries in de Lima Santos *et al.* (2024), supported in part by the European Commission.

¹³⁸ WAN-IFRA (2023).

¹³⁹ Beckett & Yaseen (2023); and see Beckett (2019); Kothari & Cruikshank (2022); Marconi (2020).

¹⁴⁰ van Dijck *et al.* (2018b).

¹⁴¹ Horowitz *et al.* (2022).

¹⁴² Schaetz *et al.* (2023), supported by the Federal Ministry of Education and Research (BMBF, Bundesministerium für Bildung und Forschung), Germany.

¹⁴³ Meese & Hurcombe (2021).

¹⁴⁴ Schirch (2021); see also the discussion on polarization in Section 4.4, Chapter 2.

¹⁴⁵ Meese & Hurcombe (2021); van Dijck & Poell (2013).

a trend that does not uniformly affect the entire industry.¹⁴⁶

Facebook/Meta's approach to news. Facebook's interest in news content has grown as it sought to monetize online advertising and counter X (then Twitter)'s emerging status as a key news source. In 2013 the company began promoting news publishers' content in its personalization system. This encouraged news organizations to focus on Facebook distribution strategies for their news. Facebook developed technologies for hosting content directly (e.g., the launch of Instant Articles), and incentivized publishers to keep their content on its platform. The platform's shift to video content and the introduction of Facebook Live led the media industry to adapt to these changes. The relationship between publishers and Facebook soured due to monetization challenges, inflated video metrics by Facebook, and controversies surrounding mis- and disinformation, especially during the 2016 United States presidential election. Facebook's response was to step away from news distribution in 2018, changing its News Feed algorithm to prioritize personal content. Faced with this challenge, some news media organizations altered their distribution strategies, aiming to regain control of revenue streams and favor core audience interests over Facebook demands.¹⁴⁷

The extent to which the push to adopt AI tools will increase news media dependency on digital platforms is unclear.¹⁴⁸ Claims that the 'AI goldrush' will increase the potential for infrastructure capture and shift even more control to platform companies raises questions about control, dependence and autonomy, as the adoption of AI tools in newsrooms extends platform control over the news production processes and the distribution networks.¹⁴⁹ While there is a growing market for AI tools to cater to newsrooms' needs, with smaller players such as Narrativa, Retresco, Adobe and others trying to position themselves in the market, the dominant players operate in an oligopolistic market (see Table 3.1).

Table 3.1

AI systems uses in the news media gatekeeping process

Production and distribution process	Use of AI systems
Access and observation	<ul style="list-style-type: none"> • Information discovery. • Audience and trends analytics; story detection. • Prompting for new ideas following from a news story.
Selection and filtering	<ul style="list-style-type: none"> • Verification, claim matching, and similarity analysis (e.g., for fact-checking). • Content and/or document categorization; analysis of datasets. • Automated collection and analysis of structured data (e.g., financial, banking, and sports data). • Coding assistance for various tasks. • Transcription and translation of audio and video. • Search in archives and/or metadata.
Processing and editing	<ul style="list-style-type: none"> • Brainstorming and ideation. • Content production (writing of draft text or articles; editing of news content). • (Re-)formatting of content for online, social media, print, broadcast (e.g., summarization, simplification, stylistic changes; text-to-video, speech-to-text, text-to-speech translation). • Copy editing, adaptation to house style. • Tagging of content, headline, and SEO [search engine optimization] suggestions.
Publishing and distribution	<ul style="list-style-type: none"> • Personalization and recommendation. • Dynamic paywalls, audience analytics. • Content moderation.

Source: Simon (2024, p. 13).

Multiple factors influence the extent of news organizations' dependence on digital platforms and their AI tools. These include the country (there is, for example, weak evidence of dependence in Germany); the kind of news organization; whether organizations are established, legacy or digital only (a study in South Korea, for example, found that legacy organizations experienced greater pressure than digital only); and how PSM addresses its public role and its relationship to audience reach (e.g., to young people).¹⁵⁰

¹⁴⁶ Bakke & Barland (2022).

¹⁴⁷ Meese & Hurcombe (2021).

¹⁴⁸ Simon (2022).

¹⁴⁹ Simon (2022).

¹⁵⁰ See (Hase *et al.*, 2023) whose findings are challenged by Eichler (2023); see also Poell *et al.* (2023); Pyo (2022); van Dijck *et al.* (2018a).

The wave of enthusiasm surrounding AI systems centers around its potential to transform the social roles of journalism, especially as it supports the profession's core functions in a democracy.¹⁵¹ Analysis of media coverage of 'AI' use in journalism over a five-year period in the United Kingdom and the United States indicates that opinions are far from uniform.¹⁵² There is a tension between the industry (newsroom leaders and funders) advocating for the use of this technology long-term, and professionals (journalists) highlighting concerns, for example, about the impact of AI systems on accuracy, fairness and transparency.¹⁵³ The use of AI systems in journalism is normatively evaluated in relation to stages of news work – information gathering, selection and production, and distribution and consumption, normative dimensions of accuracy, accessibility, diversity, relevance and timeliness.¹⁵⁴

Contributions of AI systems to fulfilling journalism's democratic role. Discussions on how to develop AI tools responsibly should be grounded in a normative perspective on the underlying values and principles, including the need to start with identification of values and principles with multiple stakeholders; the development of a forward-looking vision on the role of journalistic AI, grounded in a normative framework focused on editorial mission, fundamental rights and the democratic role of the media; and understanding how journalists, editors, managers, developers, users and other stakeholders can be empowered to become active agents in decision-making processes around the implementation of journalistic AI.

The argument for a more inclusive decision-making process comes from the realization

that AI apps are not just tools, but integral components of the public communication infrastructure, whose design is of concern to all stakeholders. The challenge is 'to design decision-making routines so that they become more accountable to the public, more inclusive and cognizant of diverse and underrepresented voices in society, and less dependent on a small number of major technology companies'.¹⁵⁵

Ethical concerns underlying the adoption of AI systems by journalists include whether automated content is consistent with editorial criteria; personalization that respects diversity and promotes a thriving public sphere; monitoring the quality of data to avoid bias; responsible safeguarding of user privacy; quality journalism with an emphasis on the human factor; funding of platforms and journalism independence; and AI systems to foster the values of journalism.¹⁵⁶ The Council of Europe's Steering Committee on Media and Information Society has published its *Guidelines for the Responsible Use of Artificial Intelligence in Journalism*, and there are numerous codes of practice to guide the use of these technologies.¹⁵⁷ There are concerns that the inclusion of AI tools in journalism routines could shift moral and editorial responsibility away from newsrooms, with consequences for public perceptions of news media bias.¹⁵⁸ A study of professionals in newsrooms in 16 countries in the Asia Pacific, Europe, Latin America, the Middle East and North Africa (MENA), North America and sub-Saharan Africa regions found that ethical concerns were significant. More than 60% of respondents were concerned about editorial quality, and many expressed a desire for AI systems transparency and the implementation of ethical guidelines.¹⁵⁹

¹⁵¹ Lin & Lewis (2022), supported by the Ministry of Science and Technology, Taiwan; Moran & Shaikh (2022); Beckett & Yaseen (2023), supported in part by Google News Initiative.

¹⁵² Moran & Shaikh (2022).

¹⁵³ Beckett & Yaseen (2023).

¹⁵⁴ Lin & Lewis (2022), supported by the Ministry of Science and Technology, Taiwan.

¹⁵⁵ See Helberger *et al.* (2022, p. 1621), supported by the European Research Council (ERC).

¹⁵⁶ Pocino (2021, p. 19).

¹⁵⁷ Council of Europe (2023).

¹⁵⁸ Calice *et al.* (2023); Moran & Shaikh (2022).

¹⁵⁹ Beckett & Yaseen (2023), supported in part by Google News Initiative.

In addition, developing AI model de-biasing techniques has been found to be very challenging for journalists in other studies, and the capacity to address this issue depends on the data quality that is available to journalists in their work.¹⁶⁰

3.4 USE OF GENERATIVE AI BY MIS- AND DISINFORMATION ACTORS

The widespread adoption of AI systems for content generation and distribution is associated with an increase in the spread of mis- and disinformation.¹⁶¹ In their response to the draft amendments to the IT rules in 2021, in India, IT for Change emphasized that ‘approaches to addressing misinformation and fake news need to be reframed with due cognizance of the information economy and its technological mechanics’.¹⁶² The accessibility and sophistication of content produced by GenAI are increasing as these tools provide creative possibilities for producing or altering textual, visual, auditory and audiovisual data, and are used by both private and state actors.¹⁶³

A survey conducted by Freedom House in 2023 found that a minimum of 47 countries employed commentators to manipulate online discussions in their favor, which is double the number of countries involved a decade ago.¹⁶⁴ As indicated, the evidence on how systematic these efforts are and which specific actors are involved is missing or weak, despite the fact that these ‘disinformation tactics’ are growing in sophistication as GenAI tools become more powerful, readily accessible and user-friendly. It is clear that they are being used to foment uncertainty, defame adversaries and sway public discourse.

Figure 3.2

Example of realistic AI-generated face using the 2020 algorithm StyleGAN2



Source: Authors of report.

The proliferation of false information, propaganda and hoaxes has grown dramatically with the spread of the internet and social media. It increased further with the use of user-friendly, GenAI tools, enabling ‘deepfake’ creators to build realistic synthetic videos, audios or images of real individuals without extensive technical expertise or substantial financial resources. For example, CounterCloud – an AI model said to produce automated disinformation that is convincing 90% of the time – is reported to be usable at a cost of less than USD 400 per month.¹⁶⁵ This illustrates the cost-effectiveness and simplicity with which significant mis- and disinformation operations can be generated (see Figure 3.2).

In the United States, AI-generated information has been used to tarnish the reputations of political rivals. In Venezuela, state-controlled media used AI-generated videos featuring fabricated news anchors from a fictitious international English language network to disseminate pro-government

¹⁶⁰ Dierickx *et al.* (2023b).

¹⁶¹ See the reports under the EU Code of Practice on Disinformation, March 2024, at <https://disinfocode.eu>.

¹⁶² Rajkumar & Ashraf (2023, p. 3), IT for Change is an independent NGO, Bengaluru, India.

¹⁶³ Bontridder & Pouillet (2021).

¹⁶⁴ Funk, Shahbaz & Vesteinsson (2023) supported by Amazon, the Dutch Ministry of Foreign Affairs, Dutch Postcode Lottery, Google, the Hurford Foundation, the Internet Society, Lilly Endowment Inc., the New York Community Trust, the US State Department Bureau of Democracy, Human Rights and Labor (DRL) and Verizon.

¹⁶⁵ Funk, Shahbaz & Vesteinsson (2023), supported as above.

propaganda. Produced by Synthesia, a company specializing in the creation of personalized deepfakes, this content was widely shared on social media platforms. In 2023, during the Nigerian elections, a modified audio recording created using GenAI was shared on social media. The recording falsely claimed to provide evidence of an opposition presidential candidate's involvement in attempts to manipulate the ballots.¹⁶⁶

Ofcom's Online Nation report in 2023 found that two-thirds of online 16- to 24-year-olds and over half of 25- to 34-year-olds in the United Kingdom were worried about the future impact of GenAI on society,¹⁶⁷ reflecting a new phase in the public's growing distrust in digital technologies.¹⁶⁸ A report by the United Nations General Assembly concluded that AI-generated mis- and disinformation could 'undermine information integrity and access to information' and 'undercut the protection, promotion and enjoyment of human rights and fundamental freedoms'.¹⁶⁹

Some researchers argue that concern about the risks of AI-enabled mis- and disinformation is exaggerated, and that it distracts attention from other issues. In one study, the authors note evidence that heavy misinformation consumption is limited to people who are already more likely to seek it out, leading them to conclude that increased information quality is unlikely to have a significant effect (see Figure 3.3).¹⁷⁰ We should be wary, however, of assuming that such conclusions apply globally.¹⁷¹ A study in sub-Saharan Africa found that people displayed a greater willingness to share mis- and disinformation compared to those in the United States.¹⁷²

The potential impact of GenAI on mis- and disinformation can occur in four categories: (1) increased quantity; (2) increased perceived quality; (3) increased personalization; and (4) accidental generation of plausible but false information.¹⁷³ Measuring the scale of AI generation and the distribution of mis- and disinformation and the impact of mis- and disinformation campaigns is challenging because of the difficulties of identifying, gathering and analyzing data that fully reflect people's day-to-day online experiences. The evidence that does exist suggests that the scale of AI generation and distribution of mis- and disinformation grew significantly in the five years to 2023.¹⁷⁴ Past empirical studies of bots, for example, have concluded that they are 'omnipresent' on social media platforms such as X (formerly Twitter),¹⁷⁵ although many are used for relatively benign purposes.¹⁷⁶ A study in 2019 identified 'cyber troop' (government or political party actors tasked with manipulating public opinion online) activity in 81 countries.¹⁷⁷

¹⁶⁶ Repucci & Slipowitz (2022) supported by Google Inc., the Hurford Foundation, Jyllands-Posten Foundations, Lilly Endowment Inc., Meta Platforms Inc., and National Endowment for Democracy; Ryan-Mosley (2023).

¹⁶⁷ Ofcom (2023d).

¹⁶⁸ Dutta & Lanvin (2023).

¹⁶⁹ UN (2024c, p. 3); see also UN (2024b). This is discussed in greater depth in Chapter 5.

¹⁷⁰ Broniatowski *et al.* (2023); one author is from the Office of the Assistant Secretary for Financial Resources (ASFR), US Department of Health and Human Services; see also Motta *et al.* (2024).

¹⁷¹ Madrid-Morales & Wasserman (2022).

¹⁷² Wasserman & Madrid-Morales (2019), supported by the National Research Foundation (NRF), South Africa.

¹⁷³ Simon *et al.* (2023).

¹⁷⁴ Funk, Shahbaz & Vesteinsson (2023).

¹⁷⁵ Keller & Klinger (2019).

¹⁷⁶ Makhortykh *et al.* (2022).

¹⁷⁷ Bradshaw & Howard (2019), supported by the European Research Council (ERC), Hewlett Foundation, Luminate and Adessium Foundation.

¹⁷⁸ As discussed in Section 4, Chapter 2 and Section 2, Chapter 5.

Figure 3.3

Deepfake image of Donald Trump generated using Stable Diffusions



Source: Authors of report.

The scale of the mis- and disinformation that is generated and amplified as a result of the use of AI systems is difficult to measure, and there is a lack of consensus as to its impact, and relatively limited evidence on its impact on trust in information and news media news.¹⁷⁸ Sources of evidence are variable in quality, level of detail and overall reliability. They include incidence reports, some corporate case studies, some surveys of worldwide campaigns, such as the Global Inventory of Organized Social Media Manipulation, and reports on content takedowns by platforms such as Facebook.¹⁷⁹ A lack of standards for, and transparency in, data collection, makes it difficult to verify and replicate findings.¹⁸⁰ The United Nations Policy Brief on information integrity on digital platforms documents several cases of

harm linked to mis- and disinformation, including violence against individuals and groups ensuing from the posting of hate speech, reduced take-up of Covid-19 vaccination programs and risks for the achievement of the Sustainable Development Goals.¹⁸¹ A report on a 2022 survey found that 70% of United Nations peacekeepers felt that mis- and disinformation was having a 'severe, critical or moderate impact on their work'.¹⁸² A synthesis of evidence from 1,300 sources (news articles, academic papers, white papers, and a range of other grey literature) found case studies of impact in over 70 countries.¹⁸³ Studies of the proliferation of mis- and disinformation during the Covid-19 pandemic concluded that there was a significant impact on vaccine uptake.¹⁸⁴ However, attempts to develop models to simulate the potential impact of mis- and disinformation face challenges even when data access becomes easier, in part, because there are substantial issues to be overcome in modeling real data, and many events in the world can affect how exchanges take place on platforms.¹⁸⁵

3.5 COUNTERING MIS- AND DISINFORMATION

The absence of robust AI content classification has enabled both state and for-profit actors to exploit the tendency of personalization systems to prioritize engagement-rich content.¹⁸⁶ Authoritarian nations are using AI systems to broaden and reinforce censorship. Research by Freedom House identified 22 countries that have enacted legislation mandating or providing incentives for internet platforms to use AI to eliminate speech on the internet the state deems undesirable;¹⁸⁷ for example, chatbots in China are programmed not to react to inquiries about Tiananmen Square. YouTube and X were required by the Indian government to restrict access to a documentary that showed the violence that occurred when Prime Minister Modi

¹⁷⁹ See Windwehr & York (2020). Facebook publishes annual transparency reports documenting its content moderation actions. These have been criticized for not disaggregating the types or including precise quantity of content removed. See Bradshaw *et al.* (2021), supported by the European Commission, European Research Council (ERC) and the Adessium Foundation, Civitates Initiative, Ford Foundation, Hewlett Foundation, Luminate, Newmark Philanthropies and Open Society Foundations. See Bradshaw *et al.* (2020) for country case studies and a global inventory of organized social media manipulation.

¹⁸⁰ Kostygina *et al.* (2023), supported by the National Cancer Institute and National Institute on Drug Abuse of the National Institutes of Health (NIH), US.

¹⁸¹ UN (2023a).

¹⁸² UN (2023a, p. 13).

¹⁸³ Bradshaw *et al.* (2020, 2021), supported by the European Commission, European Research Council (ERC) and the Ford Foundation.

¹⁸⁴ Fertmann & Kettmann (2021); Naeem *et al.* (2021); Posetti & Bontcheva (2020).

¹⁸⁵ Lamnitchi *et al.* (2023), funded by DARPA (Defense Advanced Research Projects Agency), US.

¹⁸⁶ Thomas (2022).

¹⁸⁷ Funk, Shahbaz & Vesteinsson (2023).

was Gujarat's chief minister. The Indian government has also urged technology companies to employ AI-based moderation techniques to regulate content.¹⁸⁸

Research has been funded by the European Commission to produce tools identifying mis- and disinformation, and several of these are used by professionals in their fact-checking work, but research also shows that tools on their own cannot counter the threat of mis- and disinformation.¹⁸⁹ As a 2023 study on anti-disinformation responses shows, tackling this information requires a unified effort that transcends individual stakeholders, such as governments acting through laws, and platforms acting through their terms of service.¹⁹⁰

Governments need to provide a legal framework for removing illegal content, and an accountability and transparency framework for problematic content, internal rules and algorithmic personalization systems, and these need to be enforced. Governments must also secure adequate funding for researchers and civil society to leverage data access rights. Additionally, promoting partnerships with digital platforms can help elevate verified information sources, supporting PSM and independent entities that contribute to democracy and education.

4 AI Systems and Democracy

This section addresses the reciprocal relationships between the development and deployment of AI systems and mediated public sphere(s), including how these relationships affect news media diversity and media freedom, and more generally, the

interaction between these systems and societal resilience and cohesion, social and environmental sustainability.

4.1 AI SYSTEMS AND MEDIATED PUBLIC SPHERE(S)

The use of AI systems for content governance shapes the public sphere(s) in which communication flows occur. While private communication platforms that use these technologies do not themselves directly 'censor', the design and use of content governance algorithms influences democratic discourses.¹⁹¹ Just as AI systems can contribute to more diverse information ecosystems, they can reinforce the monitoring capabilities of authoritarian states and enhance inequalities and unfair power structures through labor extractivism.¹⁹² Without negating the role of automated tools, it is important to realize that non-technology-related phenomena, such as the quality of a social security system or whether gender equality is supported, are found to be bigger factors when it comes to furthering societal cohesion and resilience. This means that, in assessing the impact of AI systems, the socio-economic and political context in which information ecosystems operate have to be taken into account, as well as the policy and regulatory situation of a country or region.¹⁹³

Notwithstanding these broader considerations, it is important to account for some of the specific influences that AI systems can have on the composition and functioning of the public sphere. AI tools used by platforms to curate content tend to favor emotionalizing content that can be used to increase engagement. This can reward social and political groups that communicate substantially through this content, or in that style.¹⁹⁴ Geographically dispersed and fringe

¹⁸⁸ Ryan-Mosley (2023).

¹⁸⁹ EC (2024b); Teyssou *et al.* (2017); Marinova *et al.* (2020), partially supported by the European Commission.

¹⁹⁰ Berger *et al.* (2023a). Measures including legislation, platform policies, fact-checking initiatives and literacy training aimed at achieving greater control over the creation and spread of mis- and disinformation are discussed in detail in Chapters 5, 6 and 7.

¹⁹¹ Elkin-Koren (2020), supported by the Israel Science Foundation.

¹⁹² Boix (2022); Adams (2022), prepared by an independent, non-partisan, African think tank.

¹⁹³ Breuer (2024), supported by the European Union Horizon 2020 program and Federal Ministry for Economic Cooperation and Development (BMZ, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung), Germany; see also Birwe (2024).

¹⁹⁴ Noble (2018).

groups can profit from easier connections through social media.¹⁹⁵ Personalization systems tend to amplify content algorithmically that emotionalizes and divides because platform business models demand engaging content.¹⁹⁶ Social media use is positively correlated with more diverse information consumption in some studies,¹⁹⁷ for example, and the use of interest histories (personalization based on previous behavior) to shape information consumption has been found to increase content diversity.¹⁹⁸ However, platforms receive ‘outsized attention and criticisms’ for being the main drivers of societal polarization, when it is also important to take account of the broader societal conditions.¹⁹⁹ Some argue that underlying societal inequality is a bigger threat to societal polarization, emphasizing that the relationship between AI systems development and societal conditions is reciprocal, but also characterized by power asymmetries.²⁰⁰

The protection of democratic values supporting the existence of the public sphere in the face of technological change is a key goal of the regulatory processes around platforms.²⁰¹ An increasing emphasis on user rights-related obligations for platforms has emerged since the early 2000s in court rulings and laws, especially in the European Union.²⁰² Power asymmetries between the Global North and Global Majority World give rise to key areas of conflict that are contributing to an ‘AI divide’. These include the increasing use of AI systems in Global Majority World countries, where there is a lack of investment in the underlying information ecosystem infrastructure and in content moderation capacities, for example, for smaller language communities and non-Global North cultures. These conditions are coupled with workforce ‘extractivism’ – the use of low-wage ‘ghost’ workers for training AI models.²⁰³

Political messaging and GenAI. Evidence of the use of GenAI for creating mis- and disinformation in political messaging is growing. This may be due partly to the increasing availability and low cost of GenAI tools whose use requires little or no technical expertise. Some evidence suggests that, while tools for detecting mis- and disinformation can do so with an accuracy of 80–90% on GenAI content created in the Global North, they are much less effective on content created in Global Majority countries because of biases in their training data. According to Sam Gregory, program director of the non-profit organization WITNESS: ‘As tools were developed, they were prioritized for particular markets’, and the data used to train the models, ‘prioritized English language – US-accented English – or faces predominant in the Western world’.²⁰⁴

Any discussion on the democratic implications of AI systems needs to include Global Majority World voices, and develop alternatives to current practices of exercising socio-economic and geopolitical power through algorithmic tools and datafication.²⁰⁵

Calling only for transparency of content governance systems that influence global information distribution processes *is not enough*. Engagement with diverging approaches to making algorithms used by communication actors more accountable is necessary.²⁰⁶ Key transparency challenges include information asymmetry, uncertainty and resourcing. This requires interdisciplinary engagement and decisions about legal rights to access information,

¹⁹⁵ Kreiss & McGregor (2023).

¹⁹⁶ Bail (2021); Settle (2018).

¹⁹⁷ Gil de Zúñiga *et al.* (2021); Möller *et al.* (2018).

¹⁹⁸ Möller *et al.* (2018).

¹⁹⁹ Kreiss & McGregor (2023).

²⁰⁰ Kreiss & McGregor (2023).

²⁰¹ Mökander *et al.* (2023), supported by AstraZeneca. Platform regulation is discussed in Section 4.3, Chapter 6.

²⁰² Katzenbach (2021), funded by the European Commission.

²⁰³ Monasterio Astobiza *et al.* (2022).

²⁰⁴ Elliott (2024).

²⁰⁵ Roche *et al.* (2023), funded by the Science Foundation Ireland (SFI), Centre for Research Training in Artificial Intelligence; Ricaurte (2022). See also Chapters 4 and 8.

²⁰⁶ Ananny & Crawford (2018).

shared decision-making about AI transparency choices, efforts to understand social and societal impacts, and adequate resourcing of transparency teams and audits.²⁰⁷ It is uncertain the extent to which the voices of those in Global Majority World regions will play a role as countries in the Global North push back on the United Nations' efforts to give countries in these regions – including China – a strong voice in AI systems governance.²⁰⁸

In the light of the challenges of algorithmic content production and distribution, media plurality and media diversity, as well as media freedom, must be protected. The existence, and plurality, of independent news media of sufficient quality is impacted by increased use of AI tools for content production which, in turn, is influenced by trends in market concentration triggered by AI systems investment.²⁰⁹ In 2018 the Council of Europe recommended that automated decision-making processes governing the distribution of online content should 'improve the effective exposure of users to the broadest possible diversity of media content online'.²¹⁰ Assessing how to measure media diversity is not a simple task, and proposals for metrics aimed at assessing initiatives to support a more diverse media environment are only a first step.²¹¹

Research that suggests AI systems use in social media has negative effects on content diversity in terms of its distribution may neglect the multidimensionality of diversity that encompasses 'topic plurality, genre' and 'plurality in tone'. Studies using the concept of 'exposure diversity' – the diversity of information users actually see – find that algorithmic personalization systems have strong positive effects on diversity. The 'element of surprise: serendipity' is an essential part of (most) of these systems. Highly personalized systems that increase the perceived relevance of specific content for users can reduce the range

of information they encounter, although increases in media and information (and AI) literacy may mitigate this effect.²¹² Regulatory approaches, discussed in Chapter 6, aim to address the need to receive data from platforms on key optimization goals of content governance systems.

4.2 AI SYSTEMS AND SOCIETAL RESILIENCE AND COHESION

Information ecosystems are connected to other societal systems, and although AI systems are only one factor in societal transformation processes, they can both challenge and enhance societal resilience and cohesion. Societal resilience refers to the ability of a society to react to, and recover from, challenges and disruptions, including short-term disruptions (e.g., armed attacks), medium-term crises (e.g., the Covid-19 pandemic) and long-term challenges (e.g., climate change).²¹³

Societal cohesion is a key contributing factor to, and predictor of, societal resilience. It refers to the capacity and extent to which a society's members cooperate and work together toward collective well-being based on shared values. Values are shared, questioned and developed through communication processes. When automated content moderation tools play an important role in information ecosystems, they can have an impact on societal cohesion and thus resilience. Being aware of the rules and practices governing mediated discourse is important for meaningful democratic participation, and increasing 'algorithmic awareness' is an important aspect of AI literacy.²¹⁴ By increasing sensitivity to the impact of AI systems on content production and distribution, societal cohesion and resilience can be better supported. Conversely, research suggests that greater societal resilience is positively correlated with resistance to mis- and disinformation.²¹⁵

²⁰⁷ Ruffo *et al.* (2023), funded by IBERIFIER (Iberian Digital Media Research and Fact-Checking Hub), European Digital Media Observatory (EDMO); Bates *et al.* (2023).

²⁰⁸ Alexander (2024) argues that the United Nations only ostensibly seeks to give Global Majority World actors a louder voice.

²⁰⁹ See Section 2, Chapter 2 for structural conditions affecting the financial sustainability of news media and their dependence on platforms that deploy AI tools.

²¹⁰ Council of Europe (2018, para. 2.5); see also Heitz *et al.* (2021).

²¹¹ Ranaivoson *et al.* (2022).

²¹² Helberger *et al.* (2018), supported by the European Research Council (ERC); Möller *et al.* (2018), supported by the European Research Council (ERC); Kreps *et al.* (2022). See Chapter 5 for a discussion of literacy.

²¹³ Berger *et al.* (2023b); Haas & Kettemann (2024); Kettemann & Lachmayer (2021); Veale *et al.* (2023).

²¹⁴ De Vivo (2023). AI literacy is discussed in detail in Chapter 5.

²¹⁵ Kertysova (2018).

Evidence also suggests that use of AI by political actors can increase the quality and speed of responses to political queries by citizens, thus leading to a higher level of engagement, with the important caveat that citizens must be helped to understand and trust how these systems are used.²¹⁶

There is contradictory evidence concerning whether automated content governance is a main driver of societal polarization, and hence a decline in social cohesion, although polarization dynamics is a key field of research.²¹⁷ A lack of digital literacy, societal vulnerability towards ‘information pollution’, and a preexisting ‘fragmentation’ of society are cited as playing more substantial roles. However, the prevalence of mis- and disinformation such as hate speech can lead vulnerable groups to withdraw from online discourses, thus decreasing social cohesion.²¹⁸ There is little evidence that automated content governance systems are the only contributor to polarization, but news diversity and media consumption practices can clearly be affected by ‘machine gatekeeping’.²¹⁹ Exposure to misinformation and partisan information also can elicit strong emotions, which, in some studies, is shown to lead to some ‘attitude polarization’, as discussed in Chapter 2.²²⁰

4.3 AI SYSTEMS AND SOCIAL SUSTAINABILITY

The integration of AI systems into the workplace is profoundly transforming labor conditions across industries. Content moderation is essential for maintaining the quality and safety of online platforms but, despite its importance, the job is often outsourced to contract workers who face unstable employment conditions.²²¹ Processes underpinning data collection, content labeling and training can contribute to harm, including

traumatization as a result of working with problematic content or training data, which affects underpaid workers in countries that lack stringent labor protection laws.²²²

The Amazon Just Walk Out or Amazon Go stores, where people could ‘enter at gate, shop and walk out’, were employing hundreds of workers in India, and this prompted the company to roll back the use of this technology in its stores.²²³ Uber, Lyft and DoorDash use AI systems and data analytics extensively to manage their operations.²²⁴ Drivers and delivery personnel working for these companies are typically classified as independent contractors rather than as employees. This means that many workers do not receive the benefits or protections associated with employment, such as health insurance, paid leave or job security. The intersection of surveillance capabilities, worker monitoring and labor conditions means that these companies’ uses of AI systems and their approaches to collecting and processing data are attracting attention due to the potential impacts on worker privacy, autonomy and rights, with data privacy concerns being raised in certain regions, including Africa.²²⁵

AI systems managing operations. In India, the door-step food delivery platform Swiggy has gamified insurance for its rider partners. Swiggy’s weekly ranking system allows workers to access health insurance depending on the number of ‘perfect’ deliveries they make. In 2021, Amazon designed a 30-day ‘Delivery Premier League’ (DPL) for its part-time workers, under the Amazon Flex program. Modeled after the flagship cricket event Indian Premier League, DPL gamifies delivering

²¹⁶ Muñoz (2023).

²¹⁷ Ruffo *et al.* (2023). See also Section 4.4, Chapter 2.

²¹⁸ Nordic Council of Ministers Secretariat (2023); Sirbu *et al.* (2019); Washington (2023); see also Breuer (2024), supported by the European Union Horizon program and Federal Ministry for Economic Cooperation and Development (BMZ, Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung), Germany.

²¹⁹ Evans *et al.* (2023); Ross Arguedas *et al.* (2022a).

²²⁰ Weismueller *et al.* (2023).

²²¹ Ahmad & Greb (2022).

²²² Veale *et al.* (2023).

²²³ AWS (n.d.).

²²⁴ Bitter (2024); Burrell (2016).

²²⁵ Abdulrauf & Dube (2024).

packages. Each hour spent on the platform collecting packages from mini-warehouses and delivering them to customers' homes constitutes a 'run' – a unit of scoring in cricket. The more hours spent delivering, the more deliveries riders accumulate, ultimately resulting in rewards such as smartphones, motorbikes, televisions and Amazon gift cards in addition to the flat 125 rupees (about USD 1.50) paid per hour.

In many Indian cities, local governments have implemented GPS-based systems to monitor sanitation workers to boost productivity and manage schedules, raising concerns about privacy and the dignity of labor. In cities including Patna and Pune, GPS devices are used to track the movements of sanitation workers. Amazon uses sophisticated systems to track the movements and productivity of warehouse workers. Workers are often required to pack hundreds of boxes per hour, and any time spent 'off-task' can lead to warnings or job termination.²²⁶

Data center energy demand escalates.

The rapid development and adoption of AI systems is leading to escalating demands on the digital infrastructure – data centers – that are essential to its progress. Goldman Sachs has predicted that by 2030, data center power energy demand will grow by 160%.²²⁸ An investigation in late 2024 suggested that the real emissions from data centers can be more than six times the officially reported values.²²⁹

Advances in chip technology can mitigate environmental impacts, offering greater computational outputs per watt of power consumed, thus a potential offset – although relatively minor – to the energy-intensive nature of extensive data operations.²³⁰ Google has used AI systems to enhance the energy efficiency of its data centers, reducing its cooling energy requirements by up to 40%.²³¹ In some countries resistance to the energy consumption of large data centers and computing resources is emerging and strengthening.²³² This ties in with the general growing demand for public participation in decisions impacting on sustainability agendas. Researchers are calling for holistic approaches to these issues, encompassing the whole lifecycle of AI systems development, including environmentally responsible innovation.²³³

4.4 AI SYSTEMS AND ENVIRONMENTAL SUSTAINABILITY

Training state-of-the-art AI models is an energy-intensive process. LLMs demand vast amounts of data and power-intensive training processes, involving complex calculations, run on thousands of high-powered graphics processing units (GPUs) over several weeks. This can lead to a sizable environmental footprint, because data centers are one of the major drivers of increases in energy demand and in greenhouse gas emissions.²²⁷

²²⁶ Bansal (2024); Bitter (2024); Christopher (2021); Nagaraj (2020).

²²⁷ iea50 (2024).

²²⁸ Goldman Sachs (2024).

²²⁹ O'Brien (2024).

²³⁰ Berthelot *et al.* (2024); Cows *et al.* (2023), supported in part by the Vodafone Institute; one author is on the Board of Directors for Noovle S.p.A., Italy.

²³¹ Burgess (2016); Google (2022).

²³² Velkova (2024); see also WEF (2024).

²³³ Brevini (2021); Makan (2023); Wu *et al.* (2022).

5 Chapter Summary

The central question addressed in this chapter is how AI systems development and use is co-evolving with the protection of internationally protected human rights and fundamental freedoms.

The term ‘AI’ entered popular discourse to describe – misleadingly – a class of digital systems that use AI technologies to perform tasks that were the preserve of human expertise. This chapter has reviewed common definitions of ‘AI’, and explained why ‘machine learning’ (ML) is a more appropriate term to describe the systems in use in digital platforms for content governance, and why ‘AI’ is hard to avoid given the degree to which it has entered common usage. With the advent of GenAI – which can generate new content in the form of text, images and video – the impact of AI systems on people’s experiences of information ecosystems as content audiences and consumers is growing.

This chapter has explained how human rights apply in the age of digital transformation and, specifically, how they can be upheld as novel AI systems are developed and applied in different societal fields, ranging from care work to content moderation, from journalism to lending decisions. Although we argue that calls for new human rights are misguided, we emphasize that certain human rights challenges arise specifically through the widespread use of automated content governance and how decisions in this area impact society-wide democratic decision-making processes. The focus in this chapter was particularly on algorithmic bias and fairness, the relationships between freedom of expression, information and the news media, and approaches to privacy protection and participatory rights, all of which are affected by developments in AI systems.

The chapter also looked in some detail at how AI systems are being used for content governance and the impacts of their use on information integrity. Our examination of how AI systems are being deployed for content governance emphasizes that no algorithm or training data set can be free of bias.

This has impacts on news media personalization systems, and it also creates new opportunities for the use of GenAI by those who generate and disseminate mis- and disinformation, as well as for the news media industry, with consequences for the public sphere.

Understanding the properties of AI systems, including how these are related to the way they are created, optimized and used, is essential if their impact is to be gauged and if regulation is to be effective.²³⁴ A stronger focus on explainability and accountability best practices for automatic content governance systems is crucial. This is because of the need to achieve greater transparency of AI-enabled decisions through improved understanding and by encouraging trust in AI-enabled decisions when it is warranted. The pace of innovation and adoption of AI systems, and especially the emergence of GenAI, is inevitably creating substantial gaps in knowledge about how these systems are incorporated into information ecosystems and with what consequences for the health of information ecosystems.

The synthesis of research in this chapter shows that:

- It is important for researchers to be specific about the technologies, such as algorithms or ML and LLMs, that are being examined; there is a proliferation of research and commentary that treats ‘AI’ as a single category, and this is unhelpful in the face of the need to respond differently to the risks of these systems. These vary substantially in terms of the risks they pose for human rights and societal processes of self-determination.
- It is essential to confront rule of law issues, and to take account of the variety of ways in which AI systems become embedded in people’s lives, which differ across countries and regions. Discussions about the contribution of AI systems to the health of information ecosystems, or its detrimental effects, need to be as inclusive as possible.
- Internationally agreed and protected human rights and fundamental freedoms are fully

²³⁴ AI systems governance is discussed in Section 4.4, Chapter 6 and Section 3.1, Chapter 7.

applicable in today's information ecosystems, but states need to ensure that their obligations to respect, protect and implement these rights are responsive to the specific challenges posed by the new actors, instruments and power relations in the age of digital transformation.

- Biases in AI systems are a consequence of biases in the (selection of) data on which they are trained. This is not inevitable, but rather the result of human rights-insensitive practices of AI systems developers.
- Focusing mainly on tweaking content governance practices and systems ignores the multi-faceted underlying causes of social discord and distrust that give rise to polarized public opinion. A focus on the 'public worthiness' of information, rather than on information 'disorder', is likely to be a more effective way to reveal the complex elements of visibility, access, reflexivity, mediation, influence and information legitimacy.
- There is substantial evidence that the use of AI systems in content governance can lead to rights violations. Content governance systems frame the conditions under which content is seen and with whom it is shared. A lack of transparent training and deployment of automated content governance tools challenges both individual and societal rights, including freedom of expression and information and privacy, as well as democratic participatory rights.
- No single content moderation technique can be acceptable to every online participant and no content moderation or content curation system is neutral or non-discriminatory. These systems are being deployed to (usually) achieve commercial ends, with some social media companies pursuing an additional, sometimes politicized, agenda, or attempting to reduce the prevalence of certain content categories, like political content. Safeguards are needed to prevent the platforms using these systems from intensifying existing societal inequalities, increasing polarization and contributing to information disorder.
- AI systems play an important role in newsrooms in content production and distribution. The personalization of news media may positively influence the diversity of news that online users engage with, but it is essential that algorithms and other AI tools are used transparently and ethically because of their impact on the integrity of information in public sphere. The impacts of these systems on efficiency and productivity in the news industry should not be assumed.
- AI systems are being used by a range of actors to generate and distribute false information, propaganda and hoaxes, but measuring the scale of mis- and disinformation and its impacts remains challenging, partly because of the need to access real data and to develop behavioral models.
- Governments need to provide legal frameworks for defining and removing illegal content as well as assuring accountability and transparency for problematic content, and internal rules and algorithmic personalization systems. The rules arising from these frameworks need to be enforced. The European Union's Digital Services Act and AI Act, and recommendations and conventions from international organizations, including UNESCO and the Council of Europe, offer examples of good practice, but their concrete impact is not yet clear.
- It is essential that research takes account of the reciprocal relationships between the development and deployment of AI systems and the evolution of information ecosystems, including the implications for mediated public sphere(s), societal resilience and cohesion, the social sustainability of labor markets and environmental sustainability.

Research is needed:

- To provide ongoing insight into the way human rights law is being interpreted and applied at the country (regional) level, to assess whether commitments to protect fundamental rights are being met.
- To develop improved understanding of the impacts of decisions throughout the AI

development chain on the health of information ecosystems. The impact of AI systems on how information is spread and amplified by platforms remains poorly understood due to a lack of data, the complexity of interlinked algorithmic personalization systems in use by major digital platforms and diverse country contexts.

- To assess whether improving data diversity, conducting regular algorithmic audits and enforcing transparency is likely to ensure that AI systems are developed and used responsibly and ethically to achieve algorithmic fairness, thus helping to mitigate their potentially harmful effects.
- To undertake detailed studies on the mechanisms of AI-driven mis- and disinformation campaigns and their impact on democratic discourses. This includes how news media organizations are responding, and which actors/organizations are involved in using AI tools to generate mis- and disinformation, for example, whether this is government actor-driven, amplified by bots or shared by private individuals.
- To study the societal impact of algorithmic design-making, including the operation of content governance tools to understand algorithmic decision-making and auditing processes, and to hold those responsible for deploying them accountable.
- To address the disparity between those who can access and effectively leverage AI systems and those who cannot, that is, the 'AI divide'. The implications of AI systems for democratic participation, especially in the Global Majority World, require further research to avoid deepening this divide. Research is also needed to identify barriers to participation by people from the Global Majority World in developing standards and practices for AI systems.

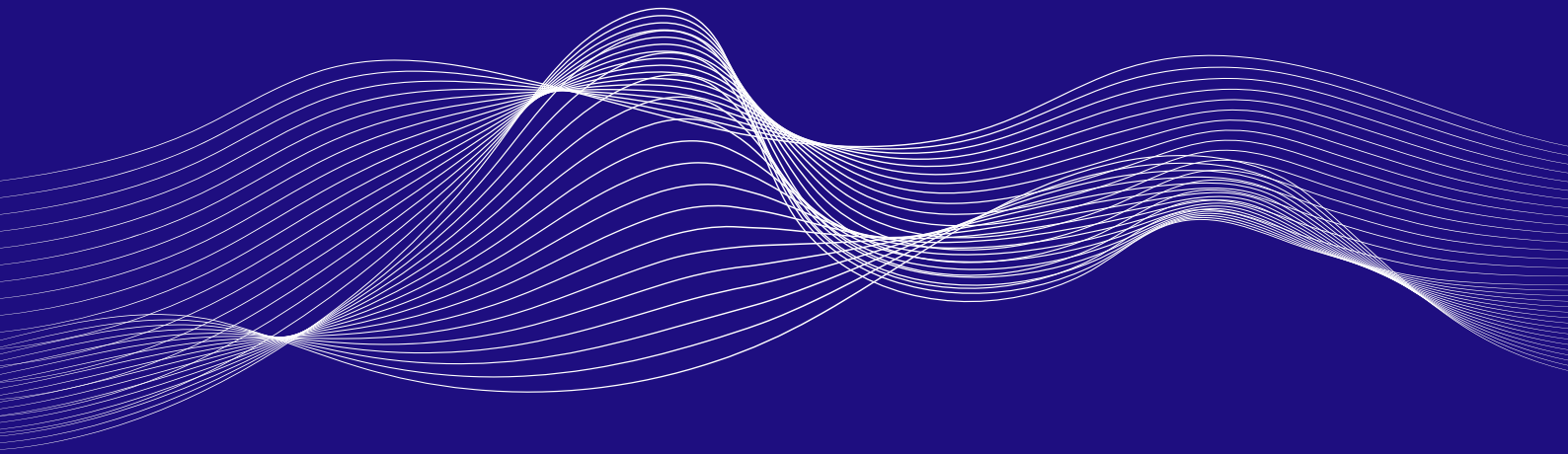


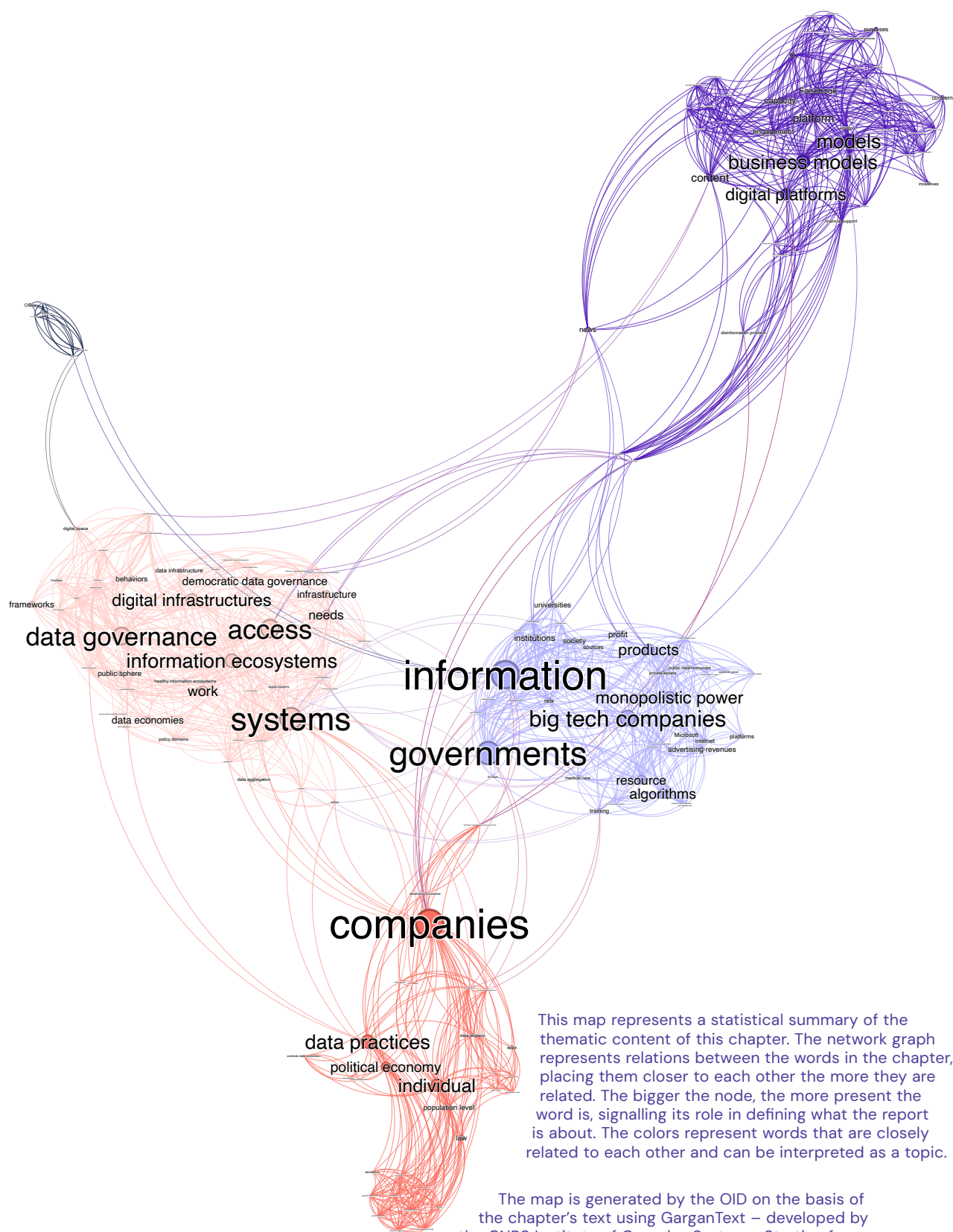
OBSERVATORY ON
INFORMATION AND
DEMOCRACY

INFORMATION ECOSYSTEMS AND TROUBLED DEMOCRACY
A Global Synthesis of the State of Knowledge on News Media, AI and Data Governance

CHAPTER 4

BIG TECH POWER AND GOVERNING USES OF DATA





[Link to the interactive map here](#)

This chapter examines evidence on the relationships between the power of big tech companies and approaches to governing the practices of data extraction and use – that is, processes of datafication.¹ It examines approaches to governing uses of data that influence data practices – the generation and the uses to which data is put – which allow information ecosystems to exist and how these practices are experienced by individuals, communities and across all sectors of economies.

The research synthesis focuses on:

- **What is the appropriate role of data and digital infrastructures within political communities?** This examines research on how to govern the massive amounts of data that are the raw material of the digital economy. Why the design and operation of data and digital infrastructures are contested is examined in the light of big tech company practices to show why these practices are inconsistent with democracy.
- **How are data aggregation and AI systems changing the way people build, share and receive information and knowledge?** This focuses on the power of big tech companies to exert monopolistic control over data through their data extraction models. It explains how this leads to data being used in ways that create dependencies of individuals, communities and industry sectors on datafication processes. It discusses how big tech business strategies result in control over information that restricts access especially for people in the Global Majority World. It highlights injustices associated with the interplay of data mining and data brokering, explaining why digital platforms are incentivized to turn a ‘blind eye’ to mis- and disinformation.
- **How do these big tech strategies and practices interfere with political deliberation which is essential for the survival of participatory democracy?** This explains why it is not sufficient to examine the harms of datafication in individualistic terms. The focus is on how data practices produce or entrench social injustices at the population level including wealth disparities and racial oppression. The need for alternative approaches to democratic data governance is discussed with a critique of measures that leave the business models and data practices of big tech companies largely in place.

This chapter provides an assessment of research in these areas providing insight into the political economy of datafication processes.

In the next chapter (Chapter 5), the scale of the mis- and disinformation problem and what the public and policy makers understand about algorithmic-driven datafication systems is discussed. Chapter 5 also examines initiatives to strengthen individuals’ capacities to control their own engagement with data-driven systems through media and information literacy as well as AI literacy. Chapters 6 and 7 discuss information ecosystem governance measures applied by governments and companies. Further discussion of data extractive practices is in Chapter 8 which critically examines alternative data governance practices.

¹ For background reading, see Aaronson (2021); Aguerre *et al.* (2024); Padovani *et al.* (2024); Taylor *et al.* (2022); Verhulst & Schüür (2023). See Appendix: Methodology for details of literature review process.

1 Introduction

The ‘information ecosystem’ is a metaphor. It is useful imagery for thinking about the dynamic, complex and interconnected set of systems that determines who can receive or share what kinds of information and in what contexts. Extending the metaphor, information ecosystems are shaped fundamentally by (and shape) the environments in which they arise – in today’s context, a data-rich environment.² An understanding of how and why mis- and disinformation arise and spread requires us to understand the political economy of data – that is, the way power relations establish the conditions for how data economies operate, and how they are inequitably experienced by different people around the world.

Data is not a naturally occurring resource, however, and analogies with natural ecosystems distort our understanding of how data aggregation and algorithmic technologies alter the systems through which people build, share and receive information. Data is produced as a result of decisions by one or more human actors to create a record of something, such as an action performed by a human, usually for a particular purpose.³ Data that constitutes today’s information ecosystems is produced and controlled primarily by a small number of companies in accordance with business models designed to prioritize profit over corporate responsibility for human rights, privacy and safety.⁴ Lurking behind the proliferation of mis- and disinformation (and other kinds of low-quality information), and the inaccessibility of useful and high-quality information, is the problem of *data governance practices* that are not designed to serve a democratically-arrived-at vision of how data should, or should not, shape the public sphere. This chapter explains why information integrity is at risk under current data governance arrangements, and why present developments in AI systems work against the requirements for healthy information ecosystems.

The OECD defines ‘data governance’ as the:

Diverse arrangements, including technical, policy, regulatory and institutional provisions, that affect data and their creation, collection, storage, use, protection, access, sharing and deletion, including across policy domains and organisational and national borders. Efforts to govern data take many forms. They often seek to maximise the benefits from data, while addressing related risks and challenges, including to rights and interests.⁵

This definition mentions both benefits and risks. It positions data governance as being as much concerned with ensuring that data is used to drive economic growth and to favor corporate interests in data monetization as it is with protecting fundamental rights. It is, however, agnostic about whether the monopolistic activities of big tech companies, including corporate practices for data capture and control, are consistent with the protection of rights. In this chapter we investigate the injustices accompanying big tech company monopolistic behaviors and consider what data governance measures are needed to ensure the uses of data generated by digital systems and applications become more closely aligned with fairness and equality.

2 Digital Infrastructure Contestations

In the 1950s Hannah Arendt worried about a near future in which human technology would replace human thinking. In *The Human Condition*, she worried not because machines would become ‘intelligent’, but because of the many things machines would make it possible for human beings to do without ‘intelligence’:

² This perspective is consistent with a socio-technical view of the interpenetration of technology and society. See Chapter 1.

³ Rosenberg (2013).

⁴ De-Lima Santos (2023), funded by the University of Amsterdam and European Union Horizon 2020. See Section 2, Chapter 2 for a discussion of digital platform company incentives in relation to the news media component of information ecosystems.

⁵ OECD (2022a, p. 13).

[I]t could be that we, who are earth-bound creatures and have begun to act as though we are dwellers of the universe, will forever be unable to understand, that is, to think and speak about the things which nevertheless we are able to do. In this case, it would be as though our brain, which constitutes the physical, material condition of our thoughts, were unable to follow what we do, so that from now on we would indeed need *artificial machines* to do our thinking and speaking. If it should turn out to be true that knowledge (in the sense of know-how) and thought have parted company for good, then we would indeed become the helpless slaves, not so much of our machines as of our know-how, thoughtless creatures at the mercy of every gadget which is technically possible, no matter how murderous it is.⁶

One way Arendt's fear has been realized is by creating infrastructures for our social, political, cultural and economic systems that are pervasive, determinative and invisible.⁷ Digital infrastructure is not disembodied; it involves familiar forms of industrial infrastructure including 'data centres distributed throughout the world and made up of servers, routers, switches, and miles of cables, as well as redundant power sources, cooling and ventilation systems, and security apparatus'.⁸ Digital infrastructure is material and just as transfiguring of physical landscapes as railroads, highways and power grids. The difference is that the digital systems that 'shape, enable and sometimes deliberately constrain life in common'⁹ are largely hidden from the conscious experience of people who depend on digital infrastructure for every aspect of life. The more that we depend on this infrastructure – to get from place to place, to shop, to access government services, to work or go to school, to get medical care and to have private conversations with friends – the more we

become part of it. The more we become part of this infrastructure, the less we are aware of it, and the more it then shapes our perceptions of everything.¹⁰

The risks of engaging with the digital environment as if it is a natural environment vary depending on the context (there may be few downsides, such as obeying traffic lights within a well-designed algorithmic road safety system).¹¹ However, the consequences of the use of digital structures and data are hugely politically significant in our systems for creating, sharing and disseminating information.¹² Much academic research and civil society advocacy examining the dysfunction of today's information ecosystems focuses on algorithmic systems. However, to think creatively about what it will take to build healthy information ecosystems, it is essential to examine the fundamental problem of how to govern the creation of, access to, and use of massive amounts of data that is the raw material of all the digital economy. Our focus here is on research that offers critical perspectives on the forces resulting in the way data is used in today's digital information ecosystems and the prospects for supporting – or even enhancing – democratic governance.

These prospects require it to be feasible for polities to contest the design of systems and the mechanisms for controlling the users of data that makes these systems possible. Digital data-dependent systems already define and constrain political discourse and activity in many contexts. For those regions of the world that have not had the opportunity for substantial input into developing today's digital economy, it is especially crucial to acknowledge parallel and often conflicting visions for the governance of both the generation and uses of data.¹³ It is also essential to recognize intersectional perspectives, including gender-sensitive approaches to data governance, and how they couple with other dimensions of

⁶ Arendt ([1958] 1998; emphasis added).

⁷ Cohen (2023); Star & Ruhleder (1996), part-funded by NSF US.

⁸ Sacasas (2021).

⁹ Stiefel *et al.* (2024).

¹⁰ Barba-Kay (2023); McLuhan (1964); Morozov (2013); Zuboff (2019).

¹¹ There may be a variety of risks associated with algorithmic traffic management systems, such as privacy infringement, especially where system maintenance depends on multilayer real-time surveillance. See, for example, Local Progress (2024).

¹² Herman & Chomsky (1989). The English word data is the plural of the Latin datum, meaning a 'thing given'.

¹³ Abdulrauf & Dube (2024).

bias or disenfranchisement.¹⁴ It is common for conversations about systems for creating and sharing information to assume that such systems must inherently rely on digital data.

The ubiquity of this concession is unsurprising since digital infrastructures are established globally and largely outside deliberative democratic processes. For example, no local, national or international legislative body considered whether it would be a good idea to set up a system whereby people access news through personalized digital filters designed to maximize the likelihood that they will eventually buy something. Digital infrastructures are imposed primarily because of under- or unregulated corporate activity alongside opaque government procurement processes.¹⁵ Many factors help to create the conditions in which the data-related features of infrastructures proliferate, typically with little political friction. None is more crucial than the lack of robust, and robustly enforced, rules about which public and private actors can do what with respect to data.¹⁶

Instances of political friction can generate political participation and change.¹⁷ Thanks to increasing public concern about corporate data practices that followed OpenAI's release of ChatGPT in late 2022 and a generative AI 'arms race', discussion around data governance issues is at an all-time high. Policy makers have historically taken up data governance in relation to the privacy, security and integrity of data, but there is strong political pressure now from within civil society to think about data governance as a lever for restructuring the markets in which technology companies operate. This is leading to efforts to protect people against infringements of their human rights, and also against concentrations of power and wealth that result in practices that are inconsistent with democracy.¹⁸

This attention gives us an opportunity to question the roles of digital data, data-dependent digital infrastructures, data markets and companies in the

data business in the very formation and function of information ecosystems. Such questioning must be part of any democratic digital policy-making project, but any such project must also seek to preserve and promote the capacities of diverse communities to take up such questioning *outside* formal policy-making spaces. This questioning is necessary *not only for* democracy, but as democracy.¹⁹

Consistent with Arendt's view, what is at stake when the data and information about the world is structured by technologies that few understand, and even fewer control is not so much the ability to resist the manipulations of technologies (as important as that may be); it is the ability to think and deliberate with others about the meaning of information and of information systems in relation to the common good.²⁰ The issues here extend far beyond protecting and promoting a healthy and inclusive public sphere, because the data practices that undergird today's information ecosystems have profound social, economic and political implications (e.g., relating to environmental impacts or wealth distribution). To explore these issues, it is necessary to understand how corporate data monopolization impacts these systems.

3 Corporate Data Monopolization and Information Infrastructures

For people living in places with a highly developed digital infrastructure, it is almost impossible to live without creating a digital record of their lives. This is increasingly so when this infrastructure starts to become more accessible to those in the

¹⁴ Chair (2024).

¹⁵ Calo & Citron (2021); Colclough (2022); Crump (2016); Hardy & Williams (2008); Zuboff (2019).

¹⁶ Cohen (2019).

¹⁷ Gordon-Tapiero *et al.* (2023); Salehi *et al.* (2015).

¹⁸ Doctorow & Giblin (2023); Mejias & Couldry (2024).

¹⁹ Benson (2019); Chambers (2023), supported by the Economic and Social Research Council UK.

²⁰ See Mazzucato (2023) for one perspective on the 'common good' as distinct from the 'public good' concept.

Global Majority World. The publication of *The Age of Surveillance Capitalism* in 2019 coincided with a surge of attention to the ways that corporations track, record and analyze our online activities to predict and shape consumers' behavior.²¹ This is the data that companies such as Google, X (formerly Twitter), Meta and Microsoft collect when people use their apps and services to create, find, consume and share information, but this is only a small part of a vast data surveillance landscape, a landscape that includes most government bureaucracies and most public and private entities that manage services and industries that are most crucial for the public.

When we do anything in a data economy, data is being produced about us: when we take public transport or drive on public roads, when we work at our jobs, when we open a bank account or apply for a credit card, when we have an interaction with a police officer or seek judicial intervention, when we apply for public benefits, when we rent or buy a home or sign up for electricity for that home, when we go to school, when we go to the doctor, when we interact in online spaces.²² As the amount of data and the number of digital repositories grow exponentially, so do the networks and digital mechanisms for sharing and selling data. Government agencies are often unaware of who has access to the data they produce about their constituents,²³ although, depending on the context, there may be rules about what the collector of data can do with it, with whom they can share it, and what a third party can then do with it.

Existing data governance frameworks consist of 'a patchwork of national regulatory regimes, multilateral bodies, corporate policies, and multi-stakeholder organizations',²⁴ and these have not proven sufficient to protect most kinds of data from being acquired by large companies that use it to generate profit or amass power.²⁵ People are being comprehensively surveilled through data production, and the

companies whose products and services shape information ecosystems are monopolizing this data. The political economy of the data infrastructure is one in which almost everyone is digitally surveilled, with the risks and burdens falling unequally on different groups and weighing most heavily on those who are already vulnerable, exploited, marginalized or targeted outside the digital context.²⁶ Those who are economically disadvantaged or subject to any form of group oppression are impacted disproportionately by the negative impacts of digital information ecosystems and by downstream misuses of data that companies commit within and for information systems, exemplified by data extraction without or with weak consent and by deploying algorithms biased in ways that benefit their economic performance. This is only compounded by the prevalence of mis- and disinformation.

The more companies achieve control of data, the more difficult it is to enact structural and systemic changes to address injustice and inequality in the digital era. There is a wide variety of corporate data practices that contribute to dysfunction and unfairness. Most involve two main types of monopolistic activity: monopolization of user data (i.e., all the data produced about us), which makes money for companies by converting information seekers into 'information products' offered for sale to advertisers; and monopolization of knowledge (i.e., data organized as usable insight) and information that makes money by converting data resources (including public data resources) into private assets. These pervasive forms of datafication give rise to numerous forms of digital dependency.

3.1 DATA MONOPOLIZATION AND DATA DEPENDENCY

Larry Page and Sergey Brin wrote an article in 1998 expressing concern about the ways that

²¹ Abdulrauf & Dube (2024).

²² On transportation, see Díaz & Levinson-Waldman (2020), supported by the Digital Industry Group Inc. (DIGI), an Australian not-for-profit industry association; on jobs, see Ajunwa *et al.* (2017); on banking, policing and court procedures, see Brayne (2020); on benefits, see Eubanks (2018); on homes and energy, see Harwell (2021); on schools, see Hooper *et al.* (2022); and on doctors, see Ledford (2019) and Corrales Compagnucci *et al.* (2022).

²³ Harwell (2019).

²⁴ LaForge & Gruver (2023).

²⁵ Mulligan & Godsiff (2023); The Majority Report w/ Sam Seder (2023); Zuboff (2019); see also Chapters 6, 7 and 8.

²⁶ Benjamin (2019); Browne (2015); Eubanks (2018); Fontes *et al.* (2022); Graham & Dittus (2022); Noble (2018); O'Neil (2016).

advertising revenue might affect the integrity of their newly launched internet search engine, Google: ‘we believe the issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm’.²⁷ Three years later, after unsuccessful attempts to sell Google to other search companies, they began selling advertising based on user data. Between 2001 and 2003 Google’s revenue increased 3,590 percent, from USD 19 million to USD 3.2 billion.²⁸ Four companies – Alphabet (Google), Meta (Facebook), Amazon, Microsoft – now largely control people’s experience of using the internet to discover and share information. The logic of the advertising attention or ‘eyeball’ economy dictates the kinds of information a person can find or receive, whether that information relates to spring fashion trends, the next election, or the history of capitalism. In the Global North, these big tech companies also dominate advertising markets – Facebook and Google together control 70 percent of the market in the United States and over 65 percent in the United Kingdom.²⁹ It is the troves of data they collect about their users in all countries in which they operate that enables them to exert such market dominance. The practices of other digital platform companies are similar even if they command smaller market shares.

These big tech companies do not limit their data collection activities to the data that they themselves extract about the people using their digital products; they also buy or license data from other companies (and acquire data analytics companies) that gather data from a wide range of public and private sources.³⁰ They also scrape and aggregate massive amounts of data from every corner of the internet. And these companies are not transparent about their uses of the data they purchase from third parties or compile from public sources, using it for targeted advertising and product development – including the training

of the algorithms that structure information flows through content moderation and curation, driving consumer activity. They sell their own data in ways that are extremely dangerous for democratic societies.³¹ If, for example, hospital systems become dependent on a managed care algorithm owned by Microsoft, the company would have significant leverage over hospital decisions about how to deliver medical care, and could make it difficult for governments to limit their data practices and data hoarding that is required to train and maintain the technology undergirding the managed care algorithm.³²

The fact that companies now take data from the internet without having to justify or compensate the data owners in any way is not the result of policy to affirmatively permit such activities. In many cases, companies determine what, if any, limits they will abide by, and they can change these at any time. Google’s 2007 terms of service read:

You give Google a perpetual, irrevocable, worldwide, royalty-free, and non-exclusive license to reproduce, adapt, modify, translate, publish, publicly perform, publicly display and distribute any Content which you submit, post or display on or through, the Services.³³

In the absence of robust data governance, tech companies treat data as an exploitable resource and, following a playbook similar to the history of European colonialism, use that data to create conditions in which resisting their continued use of that data becomes both difficult and costly.³⁴ Big tech companies use their power to amass data to reinforce their advertising dominance, squeezing out competitors, and making it difficult to develop a product or service around an alternative set of data practices, or to use, test and scale up a framework for information sharing that does not depend on advertising.

²⁷ Brin & Page (1998), funded by DARPA and NASA, and Interval Research.

²⁸ Veliz (2021, p. 32).

²⁹ Doctorow & Giblin (2023).

³⁰ Savitz (2019); The Majority Report w/ Sam Seder (2023).

³¹ See Biddle (2024), and, as in the case of Cambridge Analytica’s acquisition of data for political targeting, see Briant (2021) and Dowling (2022), supported by Department of Defense, Australia.

³² Tucker (2023).

³³ Mejias & Couldry (2024).

³⁴ Mejias & Couldry (2024).

Data monopolization is becoming more extreme as companies race to acquire the massive amounts of data needed to train algorithms to perform sophisticated classification tasks and predictive modelling. These technologies, marketed as 'AI', are extremely expensive to build because of the amount of data and computer-processing power required for training.³⁵ This means that only a small number of companies have the resources to compete in the development and training of data-intensive algorithms, and they are betting on an eventual payday that will justify their exorbitant initial investments.³⁶ The tech companies plan to integrate these technologies into their models for generating ad revenue (e.g., by using generative AI (GenAI) chatbots to mediate search activity), but it is unlikely that advertising revenue alone will yield profits that can justify the size of the bet that companies are placing on data-intensive algorithmic technology.

Every indication from big tech marketing and public relations documents is that their plan is to develop sophisticated industry-specific digital products that will offer to improve efficiency and reduce costs for companies operating within that industry, while creating 'path dependencies' that render client companies dependent on and thus 'locked in' to their algorithmic products (see Figure 4.1).³⁷ Bill Gates, whose company has invested USD 13 billion in OpenAI (that produced ChatGPT; see Figure 4.2), predicted that AI will redefine whole sectors of the economy and fundamentally change healthcare and education.³⁸

Figure 4.1
Generative AI promotion



Source: AWS Amazon.com

Figure 4.2
Large Language models for efficiency



Source: WP Event Manager

In China the government is helping to facilitate a tech oligopoly, which, at least domestically, wields economic and political power that rivals that of the largest US-based tech companies. It is doing so by strategically cultivating its national technology champions – partly by banning foreign competitors and partly by using policy incentives to favor domestic firms.³⁹ As in the United States, a number of large tech companies has emerged in China that control digital platforms for social media, e-commerce, search and online payments.⁴⁰ The country's tech industry is also expanding its geopolitical influence through the Digital Silk Road (DSR) initiative.⁴¹ This involves 'exporting' Chinese

³⁵ Mulligan & Godsiff (2023).

³⁶ Metz (2023); Novet (2023).

³⁷ OpenAI's website highlights a series of 'AI' products for sectors including healthcare and legal services (see <https://openai.com/api>). These companies are following the example of enterprise resource planning (ERP) system providers, such as SAP (Ven *et al.*, 2008); see also Ferràs-Hernández *et al.* (2023); Melih (2022); van der Vlist *et al.* (2024), supported by Dutch Research Council and German Research Foundation.

³⁸ Gates (2023).

³⁹ Tusikov (2021).

⁴⁰ Borgogno & Zangrandi (2024); Tusikov (2021); Wang (2023).

⁴¹ Erie & Streinz (2021).

information and telecommunication goods to countries (especially in Africa and the Indo-Pacific) where digital infrastructure is in the early stages of development:

... while it is true that China has consistently advocated for national autonomy over data governance issues, it is also trying to establish technological and infrastructural dependencies within the digital space of several countries. Technological dependencies are established whenever the digital infrastructure relies on standards, software and hardware that cannot be maintained without active support from Chinese players.⁴²

The problem with these data monopolization strategies is not just that a small number of companies control most of the digital data in the world or that they are using it primarily for self-enrichment.⁴³ It is also that weakly regulated corporate practices are what determined that digital data would be produced on such a scale and commercialized by default. This profoundly undemocratic, economic fiat succeeded in preempting meaningful political deliberation about rights to the ownership of digital data, what role data should have in the private and public sectors, how it should inform bureaucracy, and whether and in what contexts data production should be minimized or prohibited. Some political communities are having conversations about this now, but the terms of the debate are limited because of the entrenched data dependency globally across so many industries.⁴⁴

The cost of having ceded so much ground to tech companies is especially high when it comes to information ecosystems. When it is assumed that contemporary systems for extracting and disseminating information must operate by reducing information to ‘codable data’ and using AI tools to determine how data should flow, the questions available for people to ask about what a good

system should look like are rendered relatively one-dimensional. It may be possible for governments to compel tech companies to deliver information ecosystems that deploy algorithms that are more accurate, useful, inclusive or accessible. However, fixing the operational flaws of such algorithmic systems is ultimately a trivial problem compared with the problem for democratic society when it is not open to citizens, their representatives or the wider political community, including migrants and refugees, to contest the design and function, and even the existence, of these systems. It is this kind of thinking that involves *questioning the premise* – which is fundamental to political deliberation in democratic societies. Even the most sophisticated forms of AI do not produce systems that can question their own logics, that is, that can reflect on whether the ‘learning’ being done is the *right* learning for the problem at hand.⁴⁵

3.2 BIG TECH MONOPOLIZATION

Working in parallel with the biggest tech companies in the Global North is a cohort of lesser known, but equally powerful, companies that make money by amassing and analyzing and then selling analyzed data sets to other companies and institutions. Examples of the most powerful companies in the data analytics world in Western countries are RELX, Thomson Reuters, and Experian. These companies hoard and sell raw datasets, but also – and more importantly – information that has already been extracted from them.⁴⁶ RELX and Thomson Reuters, for example, have a duopoly with respect to legal information in the United States, owning the only two robust databases on the market for conducting legal research, both with high subscription fees. Experian is the dominant international data analytics company for financial information. It collects financial data from thousands of sources, analyzes and sells it, along with digital tools for integrating information, to businesses in more than 200 countries. There are smaller specialist data analytics companies that serve specific sectors, but

⁴² Borgogno & Zangrandi (2024, p. 19).

⁴³ Cohen (2019); Melih (2022); Mulligan & Godsiff (2023); Sadowski (2019).

⁴⁴ Rankin (2023).

⁴⁵ Green (2022); Green & Kak (2021); Selbst *et al.* (2019), supported in part by NSF and Luminate (The Omidyar Group).

⁴⁶ Lamdan (2022).

the trend is towards large data analytics companies buying others to become behemoths that own and market digital data and information products across all markets.⁴⁷

Some of the purchasers of these digital information products are other big tech companies, such as Google, and digital platforms, which incorporate the products into AI systems for search, newsfeed or ad placement. However, a large proportion of the customer base for data analytics companies is comprised of government agencies, financial institutions, law firms, universities, healthcare conglomerates and legacy media companies and news organizations. These are the entities that are typically able to pay for the products that data analytics companies sell, and they tend to function as economic, social and political gatekeepers within society. The result in many cases is that an enormous amount of information with profound public interest value is removed from the public sphere and reconstituted as the intellectual property of companies.⁴⁸ Any individual person who wants to access, for example, a newspaper article in the news archive owned by RELX containing '5 billion documents and records from over 35,000 sources of local and international news'⁴⁹ either has to be wealthy enough to afford a personal subscription, or be affiliated with an organization that has one. The consequence of placing so much information behind paywalls is often tragically concrete: 'doctors battling malaria outbreaks in Africa can't read reports about life-saving medications and measures. They can't afford to read past the articles' abstracts'.⁵⁰

The situation is especially problematic when it comes to academic research, much of which is publicly funded by taxpayers and very little of which is available to taxpayers who do not belong to an elite institution.⁵¹ Today, 'seventy-five

percent of academic research is paywalled, and it usually costs around \$30 to look at a single journal article'.⁵² In addition to creating access-to-knowledge disparities for individuals based on institutional privilege, this creates disparities among institutions. Universities in the Global Majority World are less likely to have the resources to purchase subscription services within which companies like RELX trap the academic articles they own. For instance, 'in 2008, Harvard subscribed to 98,900 serials and Yale to 73,900. The best-funded research library in India, at the Indian Institute of Science, subscribed to 10,600. Several sub-Saharan African university libraries subscribed to zero, offering their patrons access to no conventional journals except those donated by publishers'.⁵³ Scientists from all over the world recruited 27 established institutions to try to access full-text paywalled articles in the field of ophthalmology. The results showed that at 15 of those institutions researchers could access less than half the articles. Those from institutions in wealthier countries (e.g., the United States and the Netherlands) were more likely to access most of the articles. Those at institutions in Pakistan and Ecuador were unable to access any of the articles.⁵⁴ This kind of control and dominance of the information ecosystems is illustrated, for example, by Elsevier's practices.

Controlling information products: Elsevier was founded in 1880 as a publisher of scientific and medical research. It is now owned by RELX and is the dominant player in a group of five companies that control access to academic research globally. Elsevier publishes over 500,000 academic articles annually in 2,500 journals, and its archives contain over 17 million documents. With its market control, Elsevier can charge

⁴⁷ Gautier & Lamesch (2021); Lamdan (2022); Larivière *et al.* (2015).

⁴⁸ Larivière *et al.* (2015), and see Chapter 3 for a discussion of copyright and AI systems. Some government agencies do provide open access to information with the costs of acquiring and curating it borne through taxation, but they increasingly outsource to services offered by private companies.

⁴⁹ LexisNexis (2024).

⁵⁰ The Majority Report w/ Sam Seder (2023, p. 53).

⁵¹ Demeter (2019); Harvie *et al.* (2013); Nettle (2023); Puehringer *et al.* (2021).

⁵² The Majority Report w/ Sam Seder (2023, p. 53). Figures apply to the United States; charges are even higher in other parts of the world. This is changing, with many academic funders in the Global North mandating that researchers publish papers (and data where available) under open access rules. See EC (2016a); NSF (2023).

⁵³ Suber (2012, p. 30); Peters (2016).

⁵⁴ Boudry *et al.* (2019).

universities exorbitant fees for access to its journals, and package them in ways that boost the company's profit margins, rather than serving the needs of libraries and the people who use them. Elsevier makes thousands of dollars from journal articles that cost about USD 600 each to produce, generating a profit margin of 38 percent in 2023. This compares with the entire academic research industry's profit margins that hover around 30 percent, compared to Walmart's 3 percent and Toyota's 12 percent margin. After rebranding itself as an 'information analytics business', it began harvesting the data from its own content stores and using it to create digital products that do not serve the work of academics or researchers or students or librarians, but rather the money-making interests of research institutions. These products score and rank universities, journals and scholars according to prestige and influence metrics. They make predictions about which research projects will be successful. Some universities use these products to make hiring decisions, and academic funders use them to decide where to direct financial support.⁵⁵

one study put it this way: 'as the design of the algorithms and interfaces of globally dominant social media platforms maximize emotional engagement, we regard social media as a primary site of datafied emotion worldwide'.⁵⁷

'Digital influence mercenaries' exploit platform affordances on behalf of their clients, and mis- and disinformation is used to gain platform users' attention, transforming this into a commodity for sale to advertisers.⁵⁸ The introduction of the 'Like' button is a key step in the evolution of platform affordances that address the needs of influence mercenaries and platforms.⁵⁹ The 'Like' button gave Facebook a huge new source of valuable data about its users by tapping into their feelings while enhancing Facebook's personalization offering to advertisers. Facebook and other platforms are continually testing improved algorithms on its users for personalization.⁶⁰

Almost all platforms have adopted variations of these 'vanity metrics', using them to algorithmically curate content posted by other users for the purposes of identifying specific users as recipients and as targets for advertisers.⁶¹ How this content curation is performed varies among platforms. On TikTok, users' 'likes' are combined with their content views.⁶² 4chan curates users' posts so that only the most 'liked' ones survive, which leads to the promotion of more extreme material.⁶³ Platforms such as YouTube reward content producers, which is argued to incentivize the creation of more extreme content. X (formerly Twitter) has adopted a variation of this policy.⁶⁴

The question is whether these business models *inevitably* lead to platforms turning a blind eye to mis- and disinformation. After all, the reason the attention economy is key to platforms' financial viability is due to the advertising revenues that they generate on

3.3 BUSINESS MODELS AND MIS- AND DISINFORMATION

A report for the Broadband Commission, supported by the International Telecommunication Union (ITU) and UNESCO, argues that platform business models make mis- and disinformation-based campaigns attractive.⁵⁶ A core driver behind these models is an 'economics of emotion' that depends on attention and incentivizes the creation of mis- and disinformation. Focusing on the 'politics of emotion',

⁵⁵ See Nicholson (2024); The Majority Report w/ Sam Seder (2023, p. 54).

⁵⁶ Bontcheva *et al.* (2020).

⁵⁷ Bakir & McStay (2022, p. 32).

⁵⁸ See Chapter 2 for further discussion of platform business models and for a definition of affordances.

⁵⁹ Bakir & McStay (2022).

⁶⁰ For example, by so-called 'A/B' testing, where the reactions to users of two different versions of a website are tested, often without the users being aware.

⁶¹ Rogers (2018). This should not be confused with content moderation, which assesses whether a post is acceptable under a particular platform's rules.

⁶² Benton (2022).

⁶³ Tuters & Hagen (2020), supported by the European Commission.

⁶⁴ Pequeño IV (2023).

the back of it. This kind of extreme content has been found to increase user engagement: the economics of emotion monetizes deception online, first, via a service contract with digital influence mercenaries to exploit platform affordances to achieve a client's strategic objectives, and second, by attracting user attention through deceptive content and selling this attention to advertisers.⁶⁵

Having no content moderation policies (or policies that are not implemented) can be bad for business, as illustrated when advertisers terminate their business with platforms because their ads appeared alongside offensive posts such as hate speech, triggering 'brand boycotts'.⁶⁶ On balance, platform business models incentivize a very light-touch approach to content moderation and access to facilities such as their application programming interfaces (APIs).⁶⁷ X (formerly Twitter) pursues very permissive policies, allowing developers to use its APIs to create bots that automatically tweet. Facebook imposes tighter controls on the use of its APIs, but faces challenges in the management of the use of 'sock puppet' accounts (i.e., a false identity used for purposes of deception).⁶⁸

Comparing the business models of pre-digital broadcasting media, partisan media and digital media platforms reveals qualitatively different forms of mis- and disinformation. In the pre-digital media era, offending the audience was often considered bad for business, which encouraged the in some countries encouraged the presentation of news and information in ways that aligned with the beliefs and values of the majority, and with fewer tendencies towards what is today described as polarization. Regulatory changes in the United States created opportunities for the emergence of partisan media,

which led to a growth in confrontational narratives and the promotion of minority viewpoints, while partisan or state controlled media were common in other parts of the world.⁶⁹ Online digital platforms have developed more sophisticated ways of using AI systems to create platform affordances that enable the exploitation of the capacity of controversial content to capture user engagement. These platform affordances are also being exploited by a growing army of influencers who are building a following and channel content that is designed to sustain their followers' engagement, which they are then able to monetize.⁷⁰

A market-shaping approach is helpful for revealing how 'market-makers bring markets into existence through their day-to-day practices, and how their goal of generating viral content – and "clickbait" – incentivizes the circulation of "controversial claims, adversarial narratives and deceptive content"'.⁷¹ Thus, mis- and disinformation are an 'expected outcome, not breakage', of the platformized media market: far from being evidence of a dysfunctional business model, it is the outcome that is expected given these business models.

Political campaigns world-wide are increasingly data-driven as the platforms' capacities to deliver targeted advertising become more sophisticated. The political sphere is an area where platform policies regarding transparency are a particularly important concern.⁷² The Africa Center for Strategic Studies reported that mis- and disinformation campaigns increased nearly fourfold from 2022 to 2024, with a total of 189, and nearly 60 percent of these campaigns were sponsored by foreign states, with Russia, China and the United Emirates being prominent (see Figure 4.3).⁷³

⁶⁵ Bakir & McStay (2022).

⁶⁶ Zhu (2022), supported by Finnish National Agency for Education.

⁶⁷ Zammit et al (2021), supported by the Erasmus+ Strategic Partnership Program; see also Gorwa & Guilbeault (2020). APIs provide a means to programmatically interface with platform data. One example of a problem arising from this is the harvesting of millions of user profiles that are then used for targeted political advertising (Hinds et al., 2020, supported in part by the Economic and Social Research Council, UK).

⁶⁸ Wikipedia (2024).

⁶⁹ See Chapter 2 for an extensive examination of legacy and online news media.

⁷⁰ Diaz Ruiz (2023).

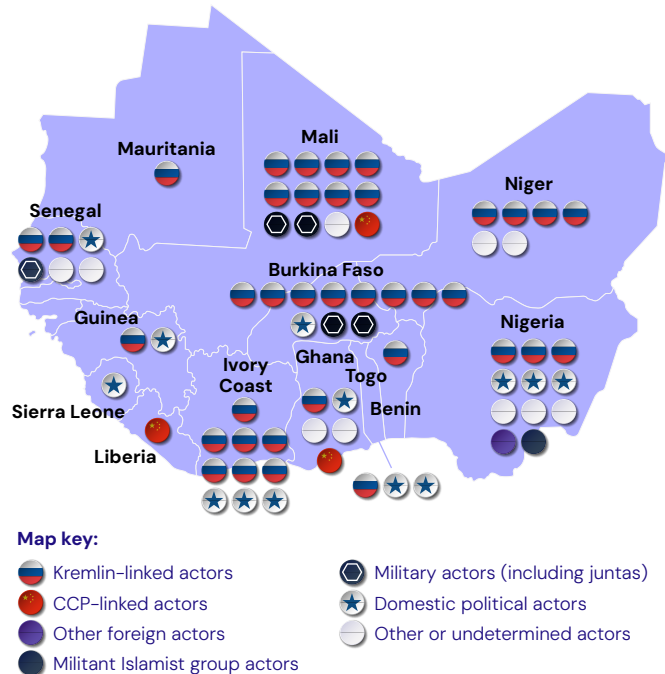
⁷¹ Diaz Ruiz (2023, p. 1).

⁷² Mehta & Erickson (2022).

⁷³ Africa Center for Strategic Studies (2024).

Figure 4.3

Map of mis- and disinformation campaigns in West Africa, and state sponsorship



Source: Africa Center for Strategic Studies (2024)

Responding to concerns about political advertising's lack of transparency, in 2018, Google, Facebook and X (formerly Twitter) established political ad archives, including information about advertisers. However, various factors, including doubts about the capacity of citizens to find and understand data and financial incentives, work against the delivery of transparency in an effective and meaningful way.⁷⁴

The United Nations argues that 'digital platforms should move away from business models that prioritize engagement above human rights, privacy and safety'.⁷⁵ A report commissioned by NATO's Strategic Communications Centre of Excellence remarks that 'buying manipulation remains cheap' and 'the gap between countering inauthentic

engagement and platform reporting is widening' as platforms focus more on limiting the reach and impact of messaging and less on denying access to commercial information manipulators.⁷⁶ Others observe that the cost of buying manipulation over time is stable,⁷⁷ raising the question of how it could be made more costly for major generators of mis- and disinformation, and what impact this might then have.

Those who are intent on spreading mis- and disinformation exploit the platforms' business models and encourage their complicity in campaign propagation.⁷⁸ For example, platforms provide a target for 'cyber troop' activity (i.e., government or political party actors tasked with manipulating public opinion online), with activity identified in at least 81 countries.⁷⁹ The data economy fosters a highly competitive labor market in datafication where people persuade themselves that if they do not take on work, others will. Anonymity is afforded to online laborers when they participate in datafication work.⁸⁰ In one experiment, 87 percent of participants were found to be willing to accept jobs involving the creation of mis- and disinformation.⁸¹

Existing data governance rules permit and even foster the amplification of mis- and disinformation through coordinated influence operations as, for example, in Venezuela. A study there revealed a range of influencer motivations, organizations, technical systems, adversaries and strategies, including recruiting and paying influencers with campaigns that were organized through hierarchies and decentralized operations. Those propagating mis- and disinformation learned continuously how to evade any defenses the platform (in this case Twitter) created.⁸² The powerful big tech companies operate in ways that are counterproductive to efforts to tackle mis- and disinformation.

⁷⁴ Mehta & Erickson (2022).

⁷⁵ UN (2023a, p. 23).

⁷⁶ Fredheim *et al.* (2023, p. 3).

⁷⁷ Bradshaw *et al.* (2021), supported by the European Research Council (ERC), Adessium Foundation, Civitates Initiative, Ford Foundation, Hewlett Foundation, Luminate, Newmark Philanthropies and Open Society Foundations.

⁷⁸ Posetti & Bontcheva (2020).

⁷⁹ Bradshaw & Howard (2019) supported by European Research Council (ERC), Adessium Foundation, Hewlett Foundation and Luminate.

⁸⁰ Shleifer (2004).

⁸¹ Cohn *et al.* (2022).

⁸² Recabaren *et al.* (2023), reporting on a semi-structured interview-based study with 19 participants. Interviews focused on: (1) incentives to contribute; (2) organizational structure; (3) resources, capabilities and limitations; (4) strategies employed; (5) operations they had participated in; (6) perception of disinformation in influence operations; (7) perception of the robustness of Twitter's defenses against influence operations activities; and (8) strategies to evade and recover from detection.

Other studies of online activity that is antithetical to healthy information ecosystems emphasize that it is crucial to examine the political economy of this activity – and not just the individual actors. For example, a study of mis- and disinformation campaigns in the Philippines and Indonesia (countries with high levels of social media activity) concluded that legacy media’s history of ownership and political collusions in postcolonial societies makes them vulnerable to narratives about ‘bias’ and ‘bigotry’. In this case, research highlighted ‘the broader (Western) discourse’ that has positioned these two countries as examples of Global Majority contexts where social media have “ruined democracy”, insofar as masses of voters are assumed to have been duped by digital disinformation campaigns’.⁸³ It is a major problem when research focuses on individuals instead of on the political economy of infrastructures and data monetization, which enables platform complicity in encouraging the villainy of mis- and disinformation actors.⁸⁴ This confirms the need for creative approaches to data governance and for democratic decision making about who should be able to make use of data.

4 Towards Democratic Data Governance

The two varieties of corporate data monopolization (to control capital and access to knowledge) work in tandem to shape information ecosystems. The tech companies that are producing and using data for directing commercial behavior are maintaining an impoverished public sphere, which serves as the default digital space in which people discover and

share information.⁸⁵ Another set of companies is capturing and cordoning off from public access rich sets of data that contain usable insights. The result is a radical difference between how an information ecosystem is experienced when operating in frameworks of institutional privilege and when operating outside such frameworks.

Where data is accessible through data repositories and networks that are designed to help people easily use this data to extract information to build knowledge, people can see themselves as readers, thinkers and participants in discourse. When most information available today is a byproduct of corporate data practices that primarily aim to satisfy goals unrelated to the social project of knowledge production, people are forced to fight a system designed to treat knowledge seeking as a consumer activity.⁸⁶ This makes it extremely difficult for people inhabiting a digital space to have experiences of shared inquiry. This dichotomy contributes to the spread of mis- and disinformation: ‘in a world where scholarly research is paywalled, it’s free to hop on YouTube to watch white supremacists spread racist theories about IQ and race. But, to read a scholarly article refuting the racist YouTubers’ baseless claims with well-researched facts, the charge is USD 37.50 to overcome Sage Publishing’s paywall’.⁸⁷ Open access does not resolve all the constraints on access to online reliable information, but it does at least reduce the cost barrier where a digital infrastructure is in place.

The comprehensive data surveillance that underlies digital infrastructure, and the way that corporations aim to monopolize surveillance data to control who has access to what kinds of information, means that the role of data in information ecosystems should not be considered solely in terms of individual impacts. The inadequacy of conceptualizing the harms of the data economy in individualistic

⁸³ Ong & Tapsell (2022, p. 252), supported by the Department of Foreign Affairs and Trade, Australia.

⁸⁴ Ong & Tapsell (2022, p. 265), footnote in quote omitted.

⁸⁵ Franks (2021).

⁸⁶ Couldry & Mejias (2019); Magalhães & Couldry (2021); Schoon *et al.* (2020); West (2019).

⁸⁷ The Majority Report w/ Sam Seder (2023, p. 74).

terms is easy to see when thinking about the way that monopolistic corporate behavior makes access to high-quality information contingent on social and economic privilege. However, a more multidimensional account of data harms is necessary to make sense of the asymmetrical structure of the markets the major tech companies are using data to create.

The data collection practices of the most powerful technology companies are aimed primarily at deriving (and producing) population level insights regarding how data subjects relate to others, not individual insights specific to the data subject. These insights can then be applied to all individuals (not just the data subject) that share these population features.⁸⁸

The most prominent approaches to data governance (including AI governance) tend to focus on protecting security (individual and/or state), property and dignity/autonomy. Robust enforcement might improve outcomes for individuals and communities in highly datafied societies, but these frameworks fail in providing a framework for contesting datafication itself.⁸⁹ A primary motivation for companies to produce and aggregate massive amounts of data is to make predictions about group membership, group characteristics and behaviors that facilitate targeting for economic or political purposes, often to enhance targeted marketing of goods and services as well as for personalizing content.

While some existing or proposed frameworks for data governance address problems of improper economic or political influence, they do not take up the underlying data practices that make such targeting possible. They do not take account of the way data practices that strive to shape individual behavior according to predictions of AI tools about group membership can produce or entrench social injustices (such as wealth disparities or racial oppression).⁹⁰ Population level data injustice is

related to, but conceptually distinct from, individual harms that people suffer as a result of facial recognition algorithms trained on data that reflects racial bias.⁹¹ A neglected problem that arises with training AI systems at the population level is that the burdens of data production are borne disproportionately by certain groups. To understand what this means concretely, consider the case of a company called Fog Data Science.

Population level data injustice: Fog Data Science is a company based in the United States founded by two former Department of Homeland Security officials. Its main product is a digital tracking program called Fog Reveal, which it sells primarily to law enforcement agencies. Police departments that subscribe to Fog Reveal have access to a database containing billions of records from 250 million mobile devices, and can conduct a variety of different searches (including search by device ID). Based on these searches, the police can develop a ‘pattern of life analysis’ – a profile of individual habits based on long-term behavioral data. Fog Data Science built and maintains the Fog Reveal database by buying domestic and international location data from data brokers, which originates from over 700 smartphone apps using a mechanism called an ‘ad ID’.

The ad ID was created as a way for advertisers to personalize offers for mobile device users. It is a random string of numbers and letters that attaches to the data that smartphone apps generate about users – bundles of data can include private information (year of birth, gender, search terms used and location). Most mobile device users do not know about ad IDs or how they work, and those who are aware that their apps are recording data about their movements rarely have a way of knowing that this data is being purchased by data brokers and

⁸⁸ Viljoen (2021, p. 577).

⁸⁹ Datafication is defined in Section 2, Chapter 1. AI system governance through legislative approaches is discussed in Section 4.4, Chapter 6 and Section 3.1, Chapter 7.

⁹⁰ Viljoen (2021).

⁹¹ Mayson (2018).

sold to the police for surveillance purposes. If an individual is aware that there is a trade-off involved in using an app, they think of it as a trade-off of their privacy for their convenience. They have no way to predict downstream uses of their data or to orient their behavior ethically with an awareness of potential downstream harms.

The fact that ad ID is being used non-transparently for policing creates risks that disproportionately impact some groups more than others. In the United States, a middle-aged White woman is much less likely to be targeted by police using Fog Reveal than a young Black man, so Black men will disproportionately experience the harms not only of Fog Reveal, but of all the companies in the chain through which Fog Reveal obtains data.⁹²

This illustrates a pervasive problem in a political economy where average levels of awareness of, and tolerance for, privacy invasion set the limits of data practices, with drastically differential real-world consequences for different groups of people. Democratic data governance requires political structures that make it possible for communities to grapple with how data production and information extraction impacts on the distribution of power among people and those entities engaging in these practices, who are already differently situated – socially, culturally and economically:

The status quo of data governance law, as well as prominent proposals for its reform ... attempt to reduce legal interests in information to individualist claims subject to individualist remedies, which are structurally incapable of representing the interests and effects of data production's population-level aims. This in turn allows significant forms of social informational harm to go unrepresented and unaddressed in how the law governs data collection, processing, and use.⁹³

Given the global reach of big tech companies, and the ease with which data is created, shared,

transferred and copied,⁹⁴ data governance interventions need to consider how population level effects of data practices manifest within, but also across, political boundaries.

5 Chapter Summary

This chapter has examined how digital data, data-dependent digital infrastructures, data markets and companies in the data business produce inequalities. It has focused on how powerful actors within social, economic and political systems determine what data is produced and how it is produced. Data aggregation techniques and the AI systems embedded in digital platforms and services are changing how people build, share and receive information and knowledge. This chapter examined research on how the business models of big tech companies contribute to the production of mis- and disinformation by creating incentives for individuals to engage in information production of this kind.

These models lead to the dependence of individuals and industry sectors on the technologies and services provided by big tech companies in data-intensive economies. The research synthesis was informed by a political economy approach which focuses on struggles to govern data practices to be consistent with people's rights and interests. This chapter has shown how the monopolistic power and data governance practices favored by big tech companies succeed in pre-empting meaningful political deliberation about issues such as rights to data ownership, what role data should have in the private and public sectors, and in what contexts data production should be minimized or prohibited.

It has emphasized that combating mis- and disinformation is a collective endeavor. It requires concerted action from governments, platform providers, civil society and political entities to question

⁹² Cyphers (2022); Greenberg (2022); Turow et al. (2023).

⁹³ Viljoen (2021, p. 578).

⁹⁴ Quintais et al. (2023).

the fairness of technology development and data uses that have unfair discriminatory consequences or do not foster information ecosystems that uphold the integrity of democratic processes.

The synthesis of research in this chapter shows that:

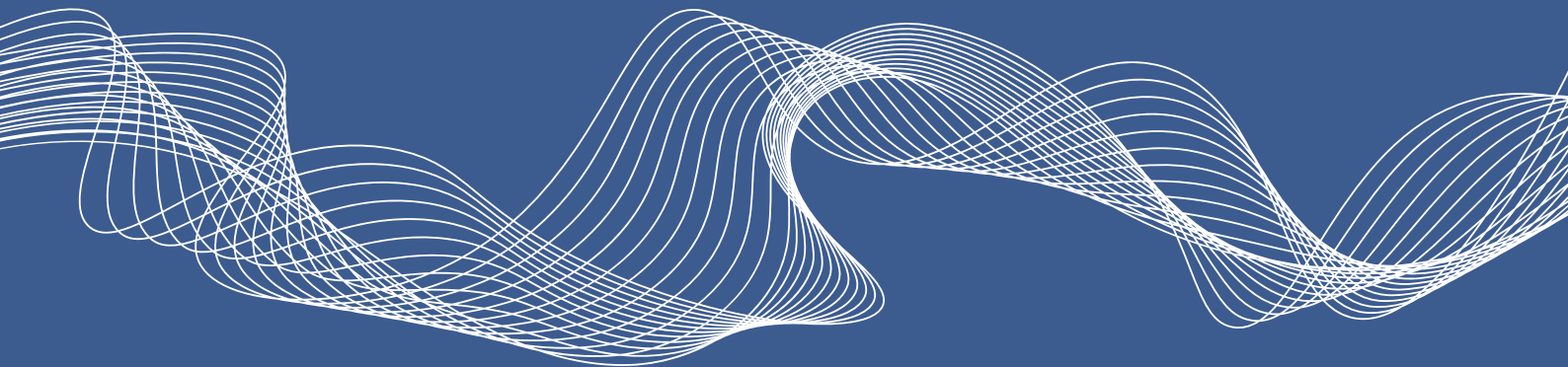
- Dependence on data-intensive algorithmic products, marketed as ‘AI’, is growing, posing significant risks to democracy. This is because when data and information are structured in ways that few understand or have control over, this affects their abilities to resist manipulations and to think and deliberate with others about the common good.
- The monopolization of data (i.e., data organized as usable insight or knowledge) occurs by converting data resources (including public data resources) into private assets. People are surveilled for data and the big tech companies do not limit their data collection to the data they extract. They buy or license data from other companies (and acquire data analytics companies) that gather or process data. Other less well known, but similarly powerful, companies also participate by amassing, analyzing and then selling data sets to other companies and institutions.
- Data governance legislation and frameworks are sufficiently permissive to foster the amplification of mis- and disinformation. These governance arrangements mean that companies and their infrastructures are creating de facto data governance frameworks that are inconsistent with data justice, and these frameworks have become normalized.
- Understanding the role of data and machine learning technologies in information ecosystems requires a multidimensional analysis of data harms that is informed by how global data dependency is becoming entrenched – that is, it must go beyond the study of impacts on individuals to focus on the political economy of power relationships and the asymmetries they produce.

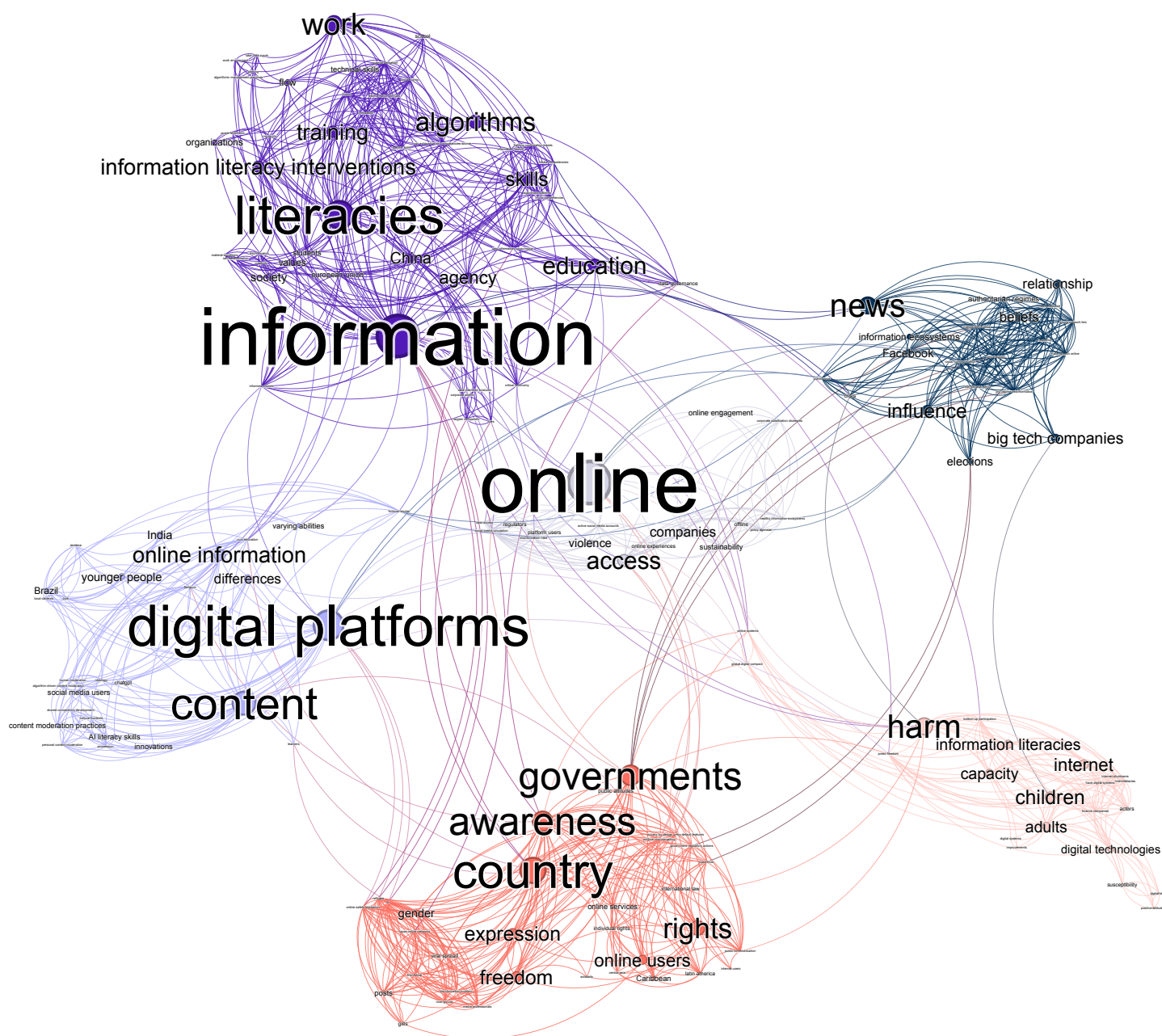
Research is needed:

- To investigate the tension between the benefits of building out network infrastructures and promoting the use of AI systems in countries in the Global Majority World where internet access is absent or very limited. Doing so risks entrenching the problems experienced in higher-income, data-intensive economies with their advanced digital infrastructures and claims to robust data governance regimes.
- To assess whether AI systems are developing in ways that are counterproductive to efforts (technical or otherwise) to tackle mis- and disinformation by investigating and exposing how big tech business models make them attractive targets for mis- and disinformation campaigns and encourage their complicity in such campaigns.
- To study how online labor markets incentivize the production of mis- and disinformation and the efficacy of steps that could be taken to discourage this.
- To investigate how extractive data production has harmful consequences for people’s daily lives, with a focus on the replication and exacerbation of inequalities and injustices.
- To examine data governance frameworks devised in countries in the Global Majority World where they are still emerging or have only recently been put in place, in order to understand what strategies are available to resist the power of big tech companies.

CHAPTER 5

AWARENESS OF MIS- AND DISINFORMATION AND THE LITERACY CHALLENGE





This map represents a statistical summary of the thematic content of this chapter. The network graph represents relations between the words in the chapter, placing them closer to each other the more they are related. The bigger the node, the more present the word is, signalling its role in defining what the report is about. The colors represent words that are closely related to each other and can be interpreted as a topic.

The map is generated by the OID on the basis of the chapter's text using GarganText – developed by the CNRS Institute of Complex Systems. Starting from a co-occurrence matrix generated from chapter's text, GarganText forms a network where words are connected if they are likely to occur together. Clustering is conducted based on the Louvain community detection method, and the visualization is generated using the Force Atlas 2 algorithm.

Link to the interactive map [here](#)

This chapter focuses on people's knowledge about the presence of mis- and disinformation in the information ecosystems they participate in, and literacy training initiatives that enable people to identify these types of information and to protect themselves from harmful consequences.¹

The research synthesis focuses on:

- **What is known about the scale and severity of mis- and disinformation?** The difficulties of measuring the production and circulation of this type of information and the extent of people's engagement with it are examined.
- **How aware are the public and policy makers of the risks and harms of mis- and disinformation?** The chapter examines whether people report concerns about mis- and disinformation generally, and in relation to political processes. Research is reviewed on awareness of how generative AI (GenAI) and algorithmic personalization systems work. Survey respondents' reports on actual or potential harms, including infringements of rights to privacy in different national contexts, are examined. Evidence on the extent of policy makers' awareness of these issues is also discussed.
- **What are the approaches to media and information literacy (MIL), and AI literacy, and what is the evidence on their effectiveness?** People's capacities to keep themselves safe online are examined, and research on MIL and AI literacy is reviewed. Evidence is examined about whether MIL and AI literacy initiatives are effective in strengthening adults' and children's agency in their online interactions, and contribute to the safer design of online systems.

This chapter provides an assessment of research in the context of the need to protect the fundamental human rights of both adults and children.

Chapters 6 and 7 discuss information ecosystems governance measures applied by governments and companies. Chapter 8 critically examines alternative data governance practices aimed at resisting injustices, biases and harms of big tech-enabled datafication practices.

¹ For background reading on literacies, see De Abreu (2022); Frau-Meigs (2024b); Frau-Meigs et al. (2017); Haider & Sundin (2022); Livingstone & Blum-Ross (2020); Ofcom (2024b); Ragnedda & Gladkova (2020); Schmarzo (2023). See Appendix: Methodology for details of literature review process.

1 Introduction

In 2023 the United Nations published a policy brief on information integrity on digital platforms. On mis- and disinformation it states: ‘the danger cannot be overstated. Social media-enabled hate speech and disinformation can lead to violence and death ... and endangers democratic institutions and fundamental human rights’.² A year later with the publication of the United Nations’ *Pact for the Future and Global Digital Compact*, the need to tackle mis- and disinformation has been linked to a broad range of societal issues, from peace to sustainability. This requires addressing:

The risks to sustaining peace posed by disinformation, misinformation, hate speech and content inciting harm, including content disseminated through digital platforms, while respecting the right to freedom of expression and to privacy and ensuring unhindered access to the Internet in accordance with international law, domestic legislation and national policies.³

The chapter begins with an overview of what is known about the scale and severity of mis- and disinformation.

2 Scale and Severity of Mis- and Disinformation

Gauging the scale and severity of the impacts of mis- and disinformation is hard. Often based on one-off studies of a single platform or limited issue-based studies, evidence indicates, for example, that

only 3% of active social media accounts produce 33% of ‘toxic’ content, or that 74% of all online conflict begins in only 1% of communities.⁴ One reason measurement is difficult is because of the problems of identifying, gathering and analyzing data that fully reflects people’s online experiences. A report prepared for Ofcom in the United Kingdom assessed the tools and methodologies that were available.⁵ Acknowledging that there is a growing range of tools and methodologies that could be applied to collect information, the report stated:

The sheer vastness and diversity of online experiences makes meaningful measurement a challenge requiring investment and innovation. The scale and variety of online platforms, and algorithmic personalisation of content, means that there is essentially *an infinite number of possible user journeys*, making it hard to arrive at both meaningful summary insights as well as fine-grained assessments of particular issues.⁶

Most of the available tools have not been designed to meet the requirements of regulators that need to gather and analyze data to better understand how to regulate the online world.⁷ In addition, gathering data on people’s online experiences poses legal and ethical issues related to privacy considerations, and most tech companies either prevent access to data or allow only selected researchers to access it. They also lobby governments to allow them to prevent data access for research or other non-commercial purposes.⁸

Studies that give an indication of the scale and severity of mis- and disinformation risks and threats come mainly from research involving samples of platform users, some following a quantitative, survey or experimental research methodology, and others, a qualitative focus group and/or interview-based approach. For example:

² UN (2023a, p. 3).

³ UN (2024b).

⁴ For a review of the literature on this type of evidence, see Robertson *et al.* (2024), supported in part by Google Jigsaw and the Templeton World Charity Foundation.

⁵ Faculty (2021).

⁶ Faculty (2021, p. 2: emphasis added).

⁷ For a review of the literature on measuring user competence in using AI tools, see Wang *et al.* (2023), supported by the National Key R&D Program of China.

⁸ Elkin-Koren *et al.* (2024).

- *Gender-based violence in the Central Sahel region.* Research in 2023 and 2024 emphasizes the need for education to tackle problems created by social media circulation of harmful content.⁹
- *Mis- and disinformation and trust in Chile.* A weak relationship was found between mis- and disinformation and media skepticism in 2017–19: initial beliefs about factually dubious information were negatively correlated with levels of trust in the news media.¹⁰ It has been argued that high levels of media trust can increase people's resilience to mis- and disinformation,¹¹ but no evidence of this was found in this study. The Chilean study was positioned as contributing to a 'nascent approach in the literature that is somewhat skeptical of the corrosive effects of misinformation on democratic regimes'.¹² This is in line with studies concluding that the corrosive effects of mis- and disinformation on attitudes toward the news media are less serious than often assumed.¹³
- *Online hunting grounds in Indonesia.* The spread of hateful content on TikTok and YouTube was seen in 2022 as receiving little pushback from authorities or the platforms.¹⁴
- *Far right presence online in Ireland.* A study in 2023 of more than 13 million posts on 12 online platforms between 2020 and 2023 found that the influence of the far right in Ireland was growing with support for White nationalism, antisemitism and Islamophobia, as well as Holocaust denial and hateful claims about the LGBTQ+ community. Big tech companies were found to be failing to enforce community guidelines, and the content on smaller platforms was found to be more extreme than on the mainstream, most-used platforms.¹⁵
- *News exposure in Mexico during the 2021 midterm elections.* A study in 2022 of the relationship between frequency of news exposure on social media platforms and beliefs in political mis- and disinformation found results consistent with a 'minimal media effects' paradigm, although platforms relying on visual communication and strong network ties were more influential.¹⁶
- *Online hate speech in the Philippines.* A study in 2022 emphasized that combating hate speech online requires broad social counternarratives and a holistic approach to tackling attacks on gendered, political and racial identities.¹⁷
- *Facebook, Russian citizens and news stories.* Two surveys in 2024 of responses to 'true' and 'false' news stories showed that the capacity of citizens living in an authoritarian regime to distinguish between them was comparable to citizens in other political contexts. Participants who mostly consumed pro-regime state media gave less accurate evaluations than those who mostly consumed independent media. Participants who were government supporters were substantially more susceptible to pro-regime misinformation than participants critical of the regime. Both pro-regime and regime critics were more likely to reject stories that were incompatible with their beliefs. 'True' critical stories were rated as false about half the time, suggesting 'that the reporting of independent media is often not a threat to authoritarian leaders'.¹⁸
- *Mis- and disinformation in six sub-Saharan Africa countries.* A study in 2022 revealed that mis- and disinformation were perceived as a problem if they had real or perceived negative consequences. Participants acknowledged a

⁹ Uyheng & Carley (2024), supported in part by the Knight Foundation and Office of Naval Research, US; see also Zullo (2023); Renaldi (2024).

¹⁰ Valenzuela *et al.* (2022), a three-wave panel study supported by the National Agency for Research and Development (ANID, Agencia Nacional de Investigación y Desarrollo), Chile.

¹¹ Humprecht *et al.* (2020).

¹² Valenzuela *et al.* (2022, p. 368), citing Allen *et al.* (2020).

¹³ Allen *et al.* (2020), supported by the Nathan Cummings Foundation, US.

¹⁴ Ong & Tapsell (2022).

¹⁵ Gallagher *et al.* (2023).

¹⁶ Valenzuela *et al.* (2024).

¹⁷ Ong & Tapsell (2022).

¹⁸ Shirikov (2024, p. 61); survey sample N = 60,000.

personal responsibility, but felt that politicians, political elites, social media platforms and governments had a greater role to play in combating the problem.¹⁹

Research on the scale and severity of the impacts of online mis- and disinformation is typically limited to a few platforms (Facebook, X/Twitter or YouTube), and is largely centered on the United States. It is essential that efforts to address mis- and disinformation go beyond the contribution that social media platforms make to the information ‘crisis’ as experienced in the United States, to take account of the varied conditions – political, social and cultural – that influence the characteristics of the information ecosystems in other parts of the world, and especially in the Global Majority World, where evidence is difficult to obtain.²⁰

There is evidence that when female journalists, for example, write reports on mis- or disinformation, digital conspiracy theories or far-right extremism, attacks increase. People engaged in producing mis- and disinformation often harass and threaten them, and this can result in their public voice and professional legitimacy being devalued.²¹ In addition, there is considerable evidence that mis- and disinformation can lead to negative impacts on public health, the quality of which depends heavily on the information available to those seeking healthcare.²²

If information about the actual scale and impact of mis- and disinformation is lacking in many parts of the world, consistent information about what people generally, and policy makers specifically, know about the factors that contribute to the likelihood that people operating within their countries will be motivated to generate such information is not systematically available across countries or over time.

3 Public and Policy Maker Awareness of Mis- and Disinformation

This section examines what is known about the public’s and policy makers’ awareness of the problems created by mis- and disinformation as an indicator of whether they are knowledgeable about what contributes to harms, and about the diverse approaches to combating these types of information in the interests of protecting fundamental rights, fostering information integrity and health information ecosystems.

With levels of ‘unfreedom’ indexed for many countries around the world being high, the public might be expected to have heightened awareness of how mis- and disinformation contribute to their ‘unfreedom’.²³ UNESCO reported in 2022 that 85% of the world’s population experienced a decline in freedom from the preceding five years.²⁴

International agreements make it clear that government measures ‘to suppress dissent and to control public communication’ must be ‘necessary and proportionate’ to protect legitimate interests under international law.²⁵ Nevertheless, as a former United Nations Special Rapporteur to the Human Rights Council observed, the potential for censorship is always present when mis- and disinformation or hate speech is circulating.²⁶ People’s views vary on whether their rights can be protected, and some research concludes that any intervention in the conduct of online interactions presents the risk of undermining freedom of expression.²⁷

¹⁹ Tully *et al.* (2022); 36 focus groups. See also Cunliffe-Jones (2021).

²⁰ Valenzuela *et al.* (2024).

²¹ Posetti *et al.* (2022).

²² Lewandowsky *et al.* (2022), citing Evanega *et al.* (2022); see also Gollwitzer *et al.* (2020); Pennycook *et al.* (2020).

²³ See country reports of the ‘unfreedom monitor’ (Advoc Team, 2024), Global Voices supported by Deutsche Welle Academy (DW Akademie) and the Federal Republic of Germany through BMZ (Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung).

²⁴ UNESCO (2022d).

²⁵ OSCE (2017, p. 2).

²⁶ Kaye (2015).

²⁷ Elsom (2020); Katsirea (2018).

3.1 PUBLIC AWARENESS OF THE IMPACTS OF MIS- AND DISINFORMATION

Research on people's awareness of risks and harms linked to mis- and disinformation varies.²⁸ A study of 142 countries found that:

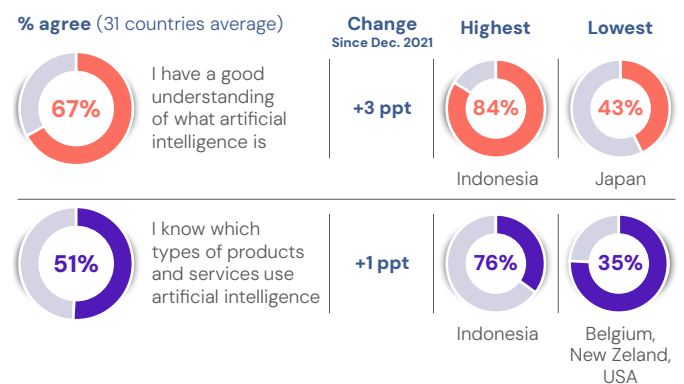
- 1) the majority of regular internet users globally (58.5%) worry about misinformation, and young and low-income groups are most likely to be concerned. 2) Risk perception among internet users *varies starkly across regions* whereby concern is highest in Latin America and the Caribbean (74.2%), and lowest in South Asia (31.2%). 3) Differences are unrelated to the prevalence of misinformation, yet concern is highest in countries with liberal democratic governments.²⁹

A survey by Ipsos and UNESCO in 2023 found that 89% of respondents agreed that 'governments and regulators should be able to require social media platforms to put in place trust and safety measures during election campaigns to protect the integrity of elections'; 85% reported being concerned about the impact of 'disinformation' in their country; and 78% reported that they read content online that had been deliberately falsified.³⁰

The social and political context and culture play a major role in public attitudes. Surveys in Mexico, South Korea, the United Kingdom and the United States found support for government regulatory actions, but this was not the case in Mexico.³¹ In the United States, online users approved efforts to increase individual choice 'rather than endorsing top-down censorship by platforms or other entities'.³² In the United Kingdom, online users reported that they would like to see more use of fact-checking.³³

Self-reported public understanding of 'AI' and automated content moderation varies considerably across countries. A global Ipsos survey in 2023 of public understanding of 'AI' in 31 countries, including the Global North and Global Majority World, indicates that overall people seem to think they have a good understanding of what 'AI' is, with fewer knowing what products and services use 'AI' (see Figure 5.1).

Figure 5.1
Understanding of AI



Source: Ipsos (2023, p.4).³⁴

Public attitudes towards the use of AI systems in the news industry also vary, as demonstrated in the case of people in Mexico, the United Kingdom and the United States, finding that comfort with the use of these tools varies on a case-by-case basis. People are generally most comfortable with GenAI being used in news production, and least comfortable with these tools being used to generate synthetic content. Disclosure about the use of AI systems does not necessarily make news consumers more trusting, and their biggest concern is about being able to detect mis- and disinformation.³⁵ A survey of the public's use of ChatGPT in Argentina,

²⁸ Public awareness of mis- and disinformation impacts is not the same as the ability to spot inaccurate information.

²⁹ Knuutila et al. (2022, p. 1; emphasis added).

³⁰ Ipsos & UNESCO (2023, p. 8). An online interview-based survey of 8,000 individuals aged 18 and over in Algeria, Austria, Bangladesh, Belgium, Croatia, Dominican Republic, El Salvador, Ghana, India, Indonesia, Mexico, Romania, Senegal, South Africa, Ukraine and the United States; a sampling of the 2.6 million total population of these countries which were scheduled for elections in 2024.

³¹ Chung & Wihbey (2024).

³² Jhaver & Zhang (2023, p. 16).

³³ Kyriakidou et al. (2023), supported by the Arts and Humanities Research Council (AHRC), UK.

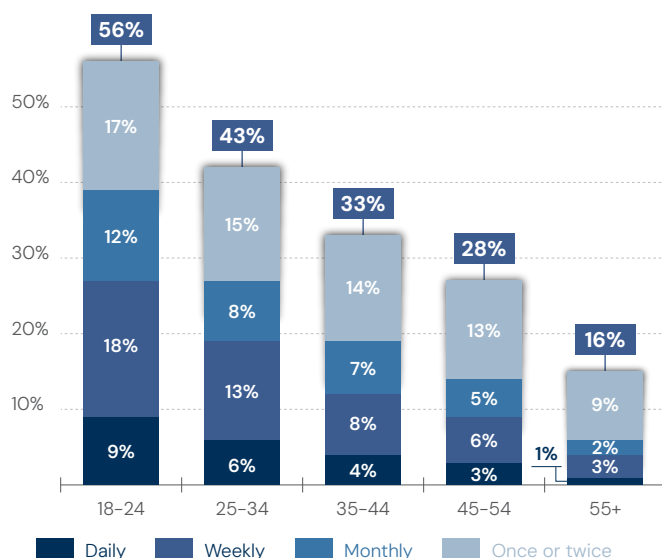
³⁴ Ipsos survey of 22,816 adults under the age of 75, with some of the samples in countries being more urban than the general population (understanding of AI ranked from highest to lowest: Indonesia, Thailand, Turkey, Malaysia, South Korea, Peru, Brazil, Romania, India, Mexico, South Africa, Chile, Singapore, Colombia, Italy, Spain, Poland, Argentina, Great Britain, Netherlands, Germany, Canada, Australia, Japan, Hungary, France, Sweden, Ireland, New Zealand, United States and Belgium).

³⁵ Collao (2024).

Denmark, France, Japan, the United Kingdom and the United States from March to April 2024 found that a plurality of people thought GenAI would make their lives better, while a significant majority believed it would worsen their lives. As far as the use of GenAI in news production is concerned, the results showed that people expect the news to be less trustworthy and transparent – and more up to date – but only a small percentage (8%) thought that the news would be worth paying more attention to if produced by AI systems.³⁶ Figure 5.2 shows the proportion of respondents who had ever used ChatGPT.

Figure 5.2

Proportion of respondents indicating they have ever used ChatGPT, by age



Source: Fletcher & Nielsen (2024, p. 12)

When individuals are asked how they would like content to be presented to them, and who or what should be responsible for decisions, they expressed varying views.

- In *Belgium*, a preference for algorithmic personalization systems that select content based on similarity was found, suggesting that the public
- do seek the means to cope with the volume of on-line information, and are also aware personalization can be performed in different ways.³⁷ People with a greater understanding of the way algorithmic personalization systems work are found to have a better understanding of the role of companies and technology developers, but this awareness is influenced by whether they experience online engagement as positive or frustrating.³⁸
- Qualitative research in *Brazil* suggests that public understanding of how algorithms work can result in political disengagement, with Brazilian Facebook users shown to stop engaging politically to avoid an 'algorithmic visibility regime' that demeans their civic voices. This might be because they believe that algorithms encourage engagement with like-minded users, that online engagement makes citizenship useless, that engagement results in unacceptable sacrifice of values and well-being, and/or success in attaining online political visibility does not mean control over visibility.³⁹
- In the *United States*, a study of marginalized youth, who depend on social media for their news and political information, found that they prefer algorithm-driven online content because they believe this enables them to exercise their agency.⁴⁰
- Another study in the *Netherlands, Portugal* and the *United States* found that algorithmic moderation was reported as being more transparent than human moderation; ironically this was particularly so when no explanation was given for the removal of content. Sending users to community guidelines on content removal had negative effects on perceptions of algorithm fairness and trust.⁴¹
- In the *United States* predictions of how people respond to algorithm-driven content moderation and selection were found to depend on the

³⁶ Fletcher & Nielsen (2024); sample sizes of around 2,000 in each country, asking 'How often, if at all, do you use each of the following AI chatbots or tools for any purpose?', showing use of ChatGPT.

³⁷ Joris *et al.* (2021).

³⁸ Martens *et al.* (2023), funded by the Research Foundation – Flanders (FWO), Belgium.

³⁹ Magalhães (2018).

⁴⁰ Kaskazi & Kitzie (2023).

⁴¹ Gonçalves *et al.* (2023), with a large sample, supported by a Facebook unrestricted gift (declared independent).

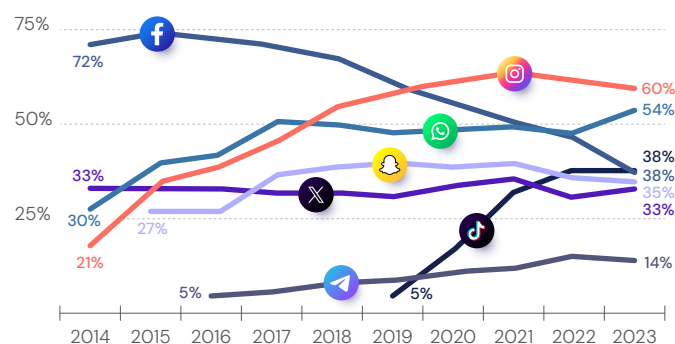
heuristic used to explain how this works, that is, whether they see machines as being more accurate and precise than humans, or believe that machines cannot make nuanced subjective judgments. These differences varied by measures of trust, fear, ideology and ability to use online services.⁴²

- Research in *Argentina, Brazil, China, Ghana, India, Jamaica, Russia, South Africa* and the countries in the *Caribbean* and *Southern African regions* confirms a mix of benefits and risks for online users in the data-driven era.⁴⁶
- In *African countries* young people's online engagement has been shown to have mixed outcomes, including uncertain long-term effects on democratic participation and evidence of relatively low levels of awareness of the impacts of social media use on young people's rights.⁴⁷
- Survey data from over 150,000 respondents in 142 countries explored perceptions of risks associated with exposure to mis- and disinformation.⁴⁸ Awareness varied: nearly 60% of regular internet and social media users registered concern about mis- and disinformation. This figure was significantly greater among people who were young and on lower incomes; people living in liberal democracies were more fearful than those living in authoritarian regimes. Concern was higher in Latin America and the Caribbean, and lower in South and Central Asia. People in some countries with a relatively high incidence of mis- and disinformation registered low levels of concern.

Figure 5.3 shows trends in the proportion of young people who use social media on a weekly basis, mainly in Global North countries.⁴³ Although increases can be noted, particularly with WhatsApp and Instagram, another study questions whether greater use leads to increased awareness of the risk of harm from online exposure.⁴⁴

Figure 5.3

Country averages of proportion of 18 to 24-year-olds using social media weekly, 2014–23



Source: Modified from Newman et al. (2023, p. 12).

People's experiences of their use of social media and their online interactions also varies considerably, particularly reflecting pre-existing beliefs and local contexts.

- Research in *Asia* indicates that people's interpretations of events and online information varies with how they decipher 'truth' in the light of local beliefs, emphasizing that this is an understudied area.⁴⁵

Views also differed about what might be done about mis- and disinformation:

- In some *African countries*, public experience of mis- and disinformation and rights is influenced in some contexts by 'ubuntu' philosophy, that is, recognizing that individual rights cannot be fully enjoyed unless the rights of all others in a community are respected. This can result in a preference for restorative justice collective measures instead of top-down government regulation.⁴⁹ This is illustrated by research on digital citizenship where an emphasis on universal

⁴² Molina & Sundar (2022).

⁴³ Those aged 18–24 in each country year in Australia, Brazil, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Spain, the United Kingdom and United States = 200. No data for Australia/Ireland in 2014. Survey question: 'Which, if any of the following, have you used for any purpose last week?'

⁴⁴ Global Kids Online (2019).

⁴⁵ Jayasinghe et al. (2022).

⁴⁶ Domingos Cordeiro & Cozman (2024); Dunn et al. (2024).

⁴⁷ Camara et al. (2023).

⁴⁸ Knuutila et al. (2022).

⁴⁹ ADRN (2024); Bayer et al. (2021, p. 74); Okyere-Manu (2023).

‘civic’ citizenship rights, at the expense of ‘ethnic conceptions of citizenship’, is found to downplay ‘hierarchies of inclusion and exclusion informed by race, ethnicity, class, gender and geography’.⁵⁰

- In *sub-Saharan Africa* evidence indicates that perceptions of the role of governments, platforms and users in stopping mis- or disinformation from circulating depends on what is believed about impacts. Qualitative research suggests that the public tends to believe that curtailing mis- and disinformation is a shared responsibility of individuals and governance institutions.⁵¹

Public awareness of how privacy infringements involving data extraction jeopardize individual rights represents a paradox, and people’s perceptions of privacy are contextual.⁵² They often claim to be concerned about privacy, but report being unwilling or unable to take steps to protect it:

- In Europe, research indicates some public awareness of how political microtargeting infringes on privacy, but also that there is uncertainty about who is responsible for data protection, the extent of excessive profiling practices and the effectiveness of privacy-by-design or by-default features of online services.⁵³
- Research in the United States demonstrates that online users are willing to trade company access to their data for ‘free’ access to platform services, even when they are concerned about unauthorized access to their data.⁵⁴
- Research indicates that ‘algorithm awareness’ is important in the decisions taken about privacy. Greater awareness and online skills influence

online users’ privacy concerns and the self-disclosure of data.⁵⁵

3.2 POLICY MAKERS’ AWARENESS OF RISKS AND HARMS

Evidence from public hearings and policy investigations suggests that policy makers may not have a sufficient understanding of how AI and platform business models operate; this may reflect in part the different inputs of technical experts and advocacy communities.⁵⁶ Campaigning and lobbying to take action against mis- and disinformation and intense media coverage can lead to ad hoc responses to risks associated with online content.⁵⁷ Ad hoc responses to events of public concern can have a chilling effect on freedom of expression, with unknown effects on the actual spread of mis- and disinformation.⁵⁸ Research in African countries shows that interventions can either restrict freedom of expression or be ineffective.⁵⁹

Viral spread of false information. In the United Kingdom, false claims posted on X that the killer of three young girls in the English town of Southport in July 2024 was a 17-year-old asylum seeker who had arrived in the country on a boat in 2023 were quickly followed by a wave of riots involving far-right groups in several cities and towns. Researchers totaled more than 27 million impressions of posts repeating this and similar false claims (see Figure 5.4).⁶⁰ According to the BBC, activity on social media platforms during this period revealed ‘a clear pattern of influencers driving a message for people to gather for protests’.⁶¹ Some argue that social media gives far-right

⁵⁰ Roberts & Bosch (2023a, p. 7) citing Nyamnjoh (2006, p. 237).

⁵¹ Tully *et al.* (2022); focus groups in six countries.

⁵² Nissenbaum (2011), supported by the Air Force Office of Scientific Research (AFSOR), US and National Science Foundation (NSF), US.

⁵³ Blasi Casagran & Vermeulen (2021), supported by Forum Transregionale Studien and Democracy Reporting International, Germany.

⁵⁴ Bright *et al.* (2022).

⁵⁵ Shin *et al.* (2022).

⁵⁶ Aula (2023), supported by the Fulbright Finland Foundation.

⁵⁷ Bunting (2018); Caplan (2023).

⁵⁸ CERTH *et al.* (2021).

⁵⁹ Cunliffe-Jones (2021).

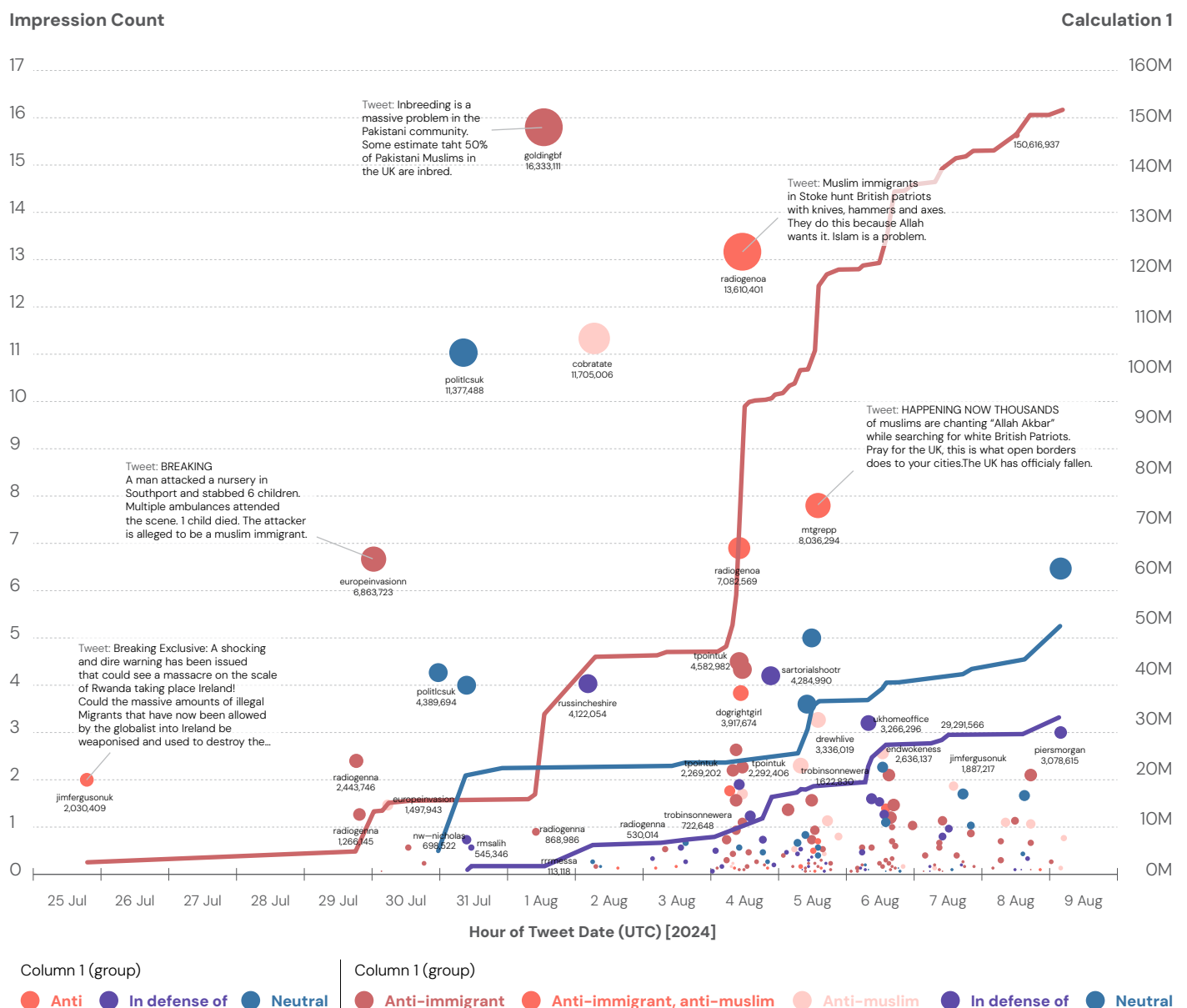
⁶⁰ Rusbridger (2024).

groups the means to extend their reach: 'All of these people are tied together by these loose online networks ... galvanised by viral online disinformation from unknown and untrustworthy sources'.⁶²

For policy makers in the United Kingdom, the immediate question was how to respond to the viral spread of false information – the person arrested and charged was not a Muslim, was not a refugee and was legally

residing in the United Kingdom. The Online Safety legislation preventing illegal online speech had not fully come into effect. Government ministers approached the owner of X, Elon Musk, to address the viral spread on false information to no effect. There were divided views about whether the X accounts promoting false information should be suppressed or taken down – some said yes, in the interests of security and safety; others, no, in the interests of protecting freedom of expression.

Figure 5.4
Spread of posts on X on Southport murders, 2024



Source: Jones (2024).⁶³

Whatever the result in this case, the deeper question for policy makers is whether they have timely evidence, the power to cause a digital platform to act, and if they do, whether it should be the state that has the power to decide what speech is amplified and what is not.

Technology and media professionals and policy makers work with diverse definitions of the mis- and disinformation problem and they develop policies in institutional silos. Their awareness of issues is conditioned by evidence that is not robust or it is contradictory or missing altogether.⁶⁴ In Global Majority World countries, as an interviewee for this report put it, 'policy makers do not understand AI or problems of mis- and disinformation, especially of marginalized or rural groups'.⁶⁵

In African countries, policy makers within state agencies may lack understanding, for example, about how to tackle mis- and disinformation and gender-related harms, and they lack the resources to assess these and other risks.⁶⁶ Even if the courts are capable of interpreting laws applicable to platform content governance systems, policies and practices, 'a lack of uniformity, limited capacity and inadequate understanding of AI means that enforcement can be erratic and uneven'.⁶⁷ In addition, in these contexts many argue that actions to address data economy challenges are 'dominated by theoretical paradigms, examples, and case studies drawn from relatively recent experiences in Global North contexts'. Without sufficient conceptual alignment, 'people end up producing distinct and incompatible things'.⁶⁸

Policy initiatives in response to mis- and disinformation are often criticized for the absence

of bottom-up participation, for failing to encourage community governance, and for neglecting the interests of marginalized groups. They may be seen as 'paternalistic' and as expecting online participants to protect themselves from risk and harm.⁶⁹ Policy makers in countries with a high level of press freedom are more likely to pursue an holistic approach to combating online mis- and disinformation, for example they are more likely to focus jointly on election processes, media and education initiatives. Countries with a higher GDP are more likely to enact legislation, while authoritarian countries are more likely to put broad legislation in place linked to penal codes.⁷⁰

Policy makers need to understand digital technologies to enable them to make sense of key terms and concepts.⁷¹ This applies as much to AI systems as it does to data and privacy protection and platform regulation. In the Southern African context, several researchers claim that 'political leaders lack understanding of what information disorder is and what impact it has'.⁷² Policy makers are charged with being too focused on the risks and threats of mis- and disinformation on social media, when a wider range of intermediaries is implicated, for example, internet domain name registries, finance companies and certificate authorities. In many countries, it is these actors that can route and address information or hack digital systems, and they have power to take down or block content or implement internet shutdowns.⁷³

One of the responses to the destabilization of democracy and to the risks and harms associated with viral mis- and disinformation is to give greater attention to initiatives designed to improve people's media and information literacy (MIL) and AI literacy.

⁶¹ Casciani & BBC Verify (2024).

⁶² Jacob Davey, Director of Policy and Research at the Institute for Strategic Dialogue (ISD), quoted in Tapper (2024).

⁶³ Figure shows which X accounts received the most impressions in the aftermath of the knife attack in Southport, United Kingdom. Red/pink dots show anti-Muslim and/or anti-immigrant tweets, brown = neutral, green = tweets defending Muslims.

⁶⁴ Carson & Wright (2022); Wasserman (2022).

⁶⁵ Interview with J. Khadijah Abdurahman, founder and Director of We Be Imagining at Columbia University's INCITE Center.

⁶⁶ RIA (2023b); Van der Spuy (2023).

⁶⁷ Hlomani (2023, p. 2).

⁶⁸ Wasserman (2022, p. 7, 112).

⁶⁹ Schneider (2022).

⁷⁰ Cipers *et al.* (2023); Ihlebæk & Sundet (2023), supported by the Research Council of Norway.

⁷¹ Mittelstadt *et al.* (2023), supported by the Wellcome Trust, Sloan Foundation, Department of Health and Social Care, UK, and Luminate Group; one of the authors worked for Amazon Web Services during part of the writing.

⁷² Sey *et al.* (2022, p. 158).

⁷³ Bradshaw & DeNardis (2022).

4 Literacies for Navigating Information Ecosystems

Evidence on people's capacities to keep themselves safe online is examined in this section, followed by a review of research on media and information literacies (MILs), and the relatively newer tradition of AI literacy, looking at whether initiatives are effective in enabling adults and children to protect themselves from threats and harms associated with mis- and disinformation.

4.1 ABILITY TO ENGAGE SAFELY ONLINE

What the public believes about the safety of online interactions varies even when they report being concerned about algorithms and platforms' data practices, but they may not have much confidence in their ability to address their concerns. In some cases, they turn for help to 'networks of literacy' (friends, colleagues, and trusted organizations) to navigate online.⁷⁴ Online users may be unable to distinguish between news personalization and platforms' involvement in targeting them for commercial reasons, but they do express concerns about their personal autonomy.⁷⁵

Children's online lives. Information about children's susceptibility to mis- and disinformation and its impact on their rights and well-being is less well developed.⁷⁶ Research on the relationship between children's digital skills and outcomes finds that better access to digital technologies at

home or at school is linked to more positive attitudes to digital technology. There is little fine-grained research on children's awareness of algorithms, the roles of digital platforms in their lives, and how their rights are affected. Studies suggest that digital skill levels are positively linked to *children's online behaviors*, for example, privacy protection, deleting messages or blocking and dealing with cyberbullying and sexual images. Skill levels are associated with whether children are likely to participate in boycotts, rallies or online campaigns, but in the case of young adults, skill levels were not found to influence whether they engage in voting and other forms of democratic participation.⁷⁷

As data-intensive technologies become more pervasive, children are impacted throughout their lives, and it is crucial to ensure that their rights are protected.⁷⁸ There are difficulties in obtaining information about the changes big tech companies make to protect the rights of children, and it is clear that changes in methods of assuring children's ages can impact on their rights to freedom of expression and non-discrimination.⁷⁹ Research does show that rights-respecting, digital design features can contribute to greater enjoyment of children when they go online.⁸⁰ However, a 10-year study of children in the United Kingdom found that 'children struggled to work out what information they could or could not trust on social media, and many were relatively unmotivated to validate the information they were seeing. Some were keen to show solidarity with views their friends had expressed, without understanding much about the issues under discussion'.⁸¹

⁷⁴ Carmi & Yates (2023); Shapiro (2019).

⁷⁵ Monzer *et al.* (2020), supported by the European Research Council (ERC).

⁷⁶ Howard *et al.* (2021).

⁷⁷ For systematic reviews of the literature, see Livingstone *et al.* (2023a); see also Livingstone *et al.* (2024).

⁷⁸ Livingstone *et al.* (2024); Mahomed *et al.* (2023).

⁷⁹ Wood (2024).

⁸⁰ Livingstone *et al.* (2023b), supported in part by the 5Rights Foundation and LEGO Group.

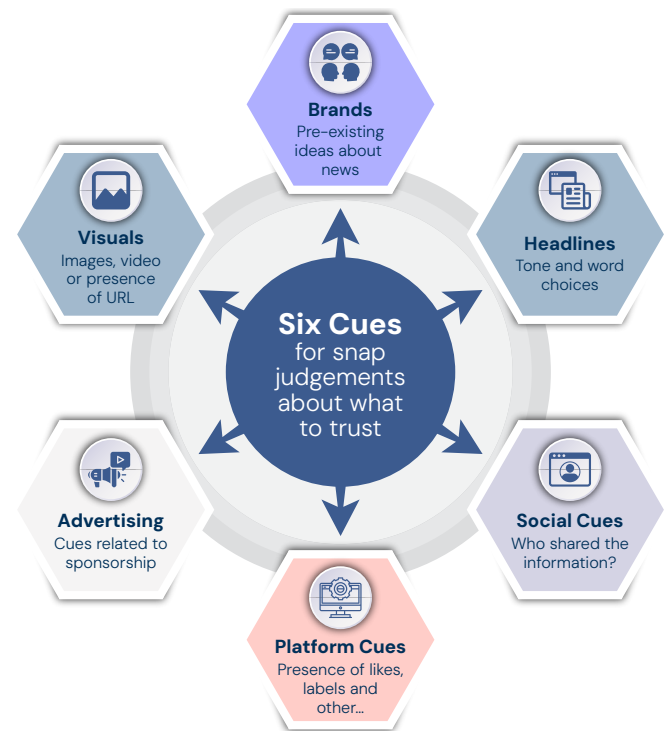
⁸¹ Ofcom (2024a, p. 7).

The *Reuters Institute Digital News Report 2023* global survey found that ‘much of the public is skeptical of the algorithms used to personalize what they see via search engines, social media, and other platforms’, suggesting relatively strong awareness about how information is managed on digital platforms, although concerns about discerning what is ‘real’ and what is ‘fake’ news varied by region.⁸² In the *United Kingdom* in 2022, 60% of social media users surveyed were confident in their ability to identify a ‘fake’ social media profile. Although 77% of users reported thinking about whether online information was truthful, there was a high risk of mistaken judgments and of being misled.⁸³ In Europe, older people generally are found to be more likely to share mis- or disinformation, echoing similar findings in *Canada* and the *United States*.⁸⁴ Research also finds that, in the case of news articles, speeches, fictional stories and recipes, people’s ability to detect whether text is authored by a human varies considerably.⁸⁵

People used a variety of competences to discriminate between trustworthy information and mis- and disinformation, but have varying abilities and competencies to do so successfully. Interviews with participants from *Brazil*, *India*, the *United Kingdom* and the *United States* investigated how people made sense of information on digital platforms (Facebook, WhatsApp and Google) and the methods they reported using to detect mis- and disinformation.⁸⁶ These methods included ‘mental shortcuts’, for example, the presence of visuals, brands, headlines and advertising sponsors. Social cues were used to assess trustworthiness, and varied with the affordances of each platform (see Figure 5.5).

Figure 5.5

Six cues for snap judgments about what to trust



Source: Ross Arguedas et al. (2022c, p. 4).

In relation to competencies to distinguish between reliable and trustworthy online content:

- A study involving *Spanish* participants explored the ‘nobody-fools-me perception’, that is, people’s overconfidence in their individual abilities to detect mis- and disinformation coupled with a self-belief that they were more immune to such information than others.⁸⁷ Younger people tended to believe that older people were more likely to be fooled by mis- and disinformation, and older people believed that younger people were less likely to fact-check. People with higher levels of education were more confident about their ability to

⁸² Newman et al. (2023, p. 10). Total sample in Africa 6,063; Latin America 12,149; Asia-Pacific 22,477; Europe 48,975. Survey question: ‘Thinking about online news, I am concerned about what is real and what is fake on the internet.’

⁸³ Ofcom (2023a).

⁸⁴ Frau-Meigs (2022); Moore & Hancock (2022); Schreurs et al. (2017), supported by the Social Sciences and Humanities Research Council (SSHRC), Canada; Jung & Sundar (2016); Hunsaker & Hargittai (2018).

⁸⁵ Dugan et al. (2023), supported in part by the Defense Advanced Research Projects Agency (DARPA), Office of the Director of National Intelligence (ODNI) and National Science Foundation (NSF), US.

⁸⁶ Ross Arguedas et al. (2022); a sample of 100 interviewees who lacked trust in their countries’ news organizations were selected.

⁸⁷ Martínez-Costa et al. (2023), funded by the BBVA Foundation and European Commission, drawing on concepts of self-perception, self-efficacy, confirmation bias, miscalibration, misplacement and mis-estimation from psychology and economics, and focusing on mis- and disinformation relating to the COVID-19 epidemic.

detect mis- and disinformation. These findings were consistent with studies suggesting that people's perception of immunity to deception tended to increase during periods of high 'information exposure',⁸⁸ as, for example, during the Covid-19 epidemic.

- A study in the *United Kingdom* in 2022 revealed that one-third of internet users were unaware of the likelihood of finding inaccurate or biased information, and some 30% did not know or think about whether the information they encountered was truthful.⁸⁹ Research in 2023 found differences in people's abilities to distinguish between different forms of online advertising, but that social media platform users were more confident in their abilities than search engine users.⁹⁰

Varying abilities of online participants to identify mis- and disinformation and to respond to it in ways that protect them from harm and enable them to participate in public debate in an informed way based on accurate information indicates that greater attention needs to be given to improving people's ability to critically assess and interpret the information they engage with. Most of the largest companies investing in data-intensive products and services are based in the United States and China. This means that in most countries and regions the powers of policy makers – apart from imposing fines or shutting services down – to force corporate actors to change how they operate are relatively weak.⁹¹

The next section examines measures to improve literacy. This is an increasingly attractive policy option, although it has been on the agenda in some countries for decades. Literacy initiatives seem to be garnering greater attention now that they are coupled with measures to improve AI literacy.

4.2 MEDIA AND INFORMATION LITERACY

Media and information literacy (MIL) initiatives aim to empower people to manage their online activity, patterns of information consumption and capacity to identify and protect themselves from harms linked to mis- and disinformation.⁹² Many MIL initiatives focus on providing people with the competences and skills to navigate their way through information ecosystems in which the integrity of information varies enormously, as do the material conditions of people's lives, including exclusions and discrimination.⁹³

MIL training is expected to:

Help realise human rights and facilitate human flourishing – including diverse forms of creativity, human connection, community and political participation – the institutions and structures of our society must make room for people's agency, knowledge and self-determination, finding ways to recognise and value and enable these, perhaps transforming themselves in the process. In short, media literacy is not a stand-alone project. To see the positives of media literacy, we have to imagine a positive vision of society – what it could be, what people want it to be, what they need it to be.⁹⁴

It is also essential to recognize that some approaches to literacy training can result in improved skills to produce and circulate hateful, dehumanizing and violent content.⁹⁵

Some warn that literacy training cannot compensate for a failure to ensure that the design of technologies is safe before services are deployed in the market.

⁸⁸ Tang *et al.* (2021). Another study of people in Spain over the age of 50 found that they were more likely to be critical of news if a headline was biased against their beliefs, and that progressive political positions were associated with greater accuracy in identifying misinformation; see Sádaba *et al.* (2023), supported in part by Meta.

⁸⁹ Ofcom (2022).

⁹⁰ Ofcom (2023a).

⁹¹ Policy and legislative measures are addressed in Chapters 6 and 7.

⁹² Depending on scientific discipline, AI literacy is either part of MIL or it is treated separately.

⁹³ For a resource on the future of media education, see Friesem *et al.* (2022) and also UNESCO's website, at www.unesco.org/en/media-information-literacy.

⁹⁴ Livingstone (2023, np).

⁹⁵ Banaji & Bhat (2022).

Education and literacy initiatives should come before experimentation and deployment of new technologies. Having so far largely failed with social media interventions and now dealing miserably with mis- and disinformation on existing social media this should not happen as AI is rolled out.⁹⁶

National, supranational and civic initiatives are underway to enhance skills to better position people to make informed choices when they engage with information online and when they share their data. States often have specific obligations to provide for or encourage MIL initiatives, particularly in the European Union and the United Kingdom,⁹⁷ but the attention and resources states devote vary remarkably in the Global North and the Global Majority World.

Definitions of MIL change over time, and research methodologies for assessing the impact of MIL initiatives differ, with numerous toolboxes developed for combating mis- and disinformation.⁹⁸ Media literacy typically refers 'to the ability to use, understand and create media and communications in a variety of contexts'. Information literacy generally refers to the ability to find, evaluate, and proficiently use information. Some refer to 'digital literacy' in relation to digital equity and inclusion.⁹⁹

Media literacy can also be defined as 'the ability to access, analyze, and produce information', the fundamental objective being 'critical autonomy in relationship to all media'.¹⁰⁰ Research increasingly focuses on conceptual frameworks that go beyond the kinds of literacy appropriate in a 'mass media' era to acknowledge the essential role of literacy in influencing how people construct their identities

and realities through their online interactions and relationships.¹⁰¹ Definitions are also being updated to reflect new patterns of media and information presentation and consumption.¹⁰²

Updating MIL definitions. Examples of definition updating come from Sri Lanka and China.

In Sri Lanka in 2021, 57.2% of those aged 5–69 were deemed to be digitally literate. Research indicated the need to revise the definition of literacy to capture not only 'computer' literacy but also literacy in the use of smartphones, the primary way that people at the 'bottom of the pyramid' access the internet. Working with the Sri Lankan ICT Agency, the definition of literacy was revised to align with UNESCO's global standard.¹⁰³

China's Education Informatization 2.0 Action Plan emphasizes 'wisdom education', shifting its investment from training principally in information technology skills towards improving student and teacher information literacy, with a focus on all-round human development, and acknowledging the need for this as datafication and AI systems become widespread.

MIL interventions aim to equip people with key competences, for example, to understand the importance of information and the ability to think critically when engaging with information and to find reliable information (see Figure 5.6).¹⁰⁴

⁹⁶ Comment by a Steering Committee member for this report.

⁹⁷ Durach *et al.* (2024); EC (2017); Frau-Meigs & Corbu (2024); Pentney (2024). In the United Kingdom, Ofcom has had media literacy duties since 2003, which were set out in the *Communications Act 2003* (UK Government, 2003) and clarified in the *Online Safety Act 2023* (UK Government, 2023), which came into effect in August 2024.

⁹⁸ Kozyreva *et al.* (2024), funded in part by Humboldt Foundation, Volkswagen Foundation, European Union Horizon program, European Research Council (ERC), Australian Research Council (ARC) and Agence nationale de la recherche (ANR), France; da Silva *et al.* (2019); Dadakhonov (2024), supported by the 'El-yurt umidi' Foundation of Uzbekistan.

⁹⁹ De Paor & Heravi (2020); Menon (2017, 2024); Ofcom (2023b, p. 3).

¹⁰⁰ Aufderheide (1993, p. 1).

¹⁰¹ For an example of new frameworks, see Cho *et al.* (2024), supported by the National Institutes of Health (NIH), US.

¹⁰² Wuyckens *et al.* (2022).

¹⁰³ Fonseka (2024), research by LIRNEasia, an independent research institute; see also UNESCO (2022a).

¹⁰⁴ Grizzle *et al.* (2021); Jones-Jang *et al.* (2021); Vuorikari *et al.* (2022).

Figure 5.6
Non-exhaustive MIL competencies



Source: Frau-Meigs (2024c, p. 4).

Attention may focus on helping online users to acquire fact-checking competences. This may involve lateral reading skills or taking independent steps to verify information. MIL initiatives may be implemented through school curricula or online (e.g., online videos, pop-ups, online games). Some work on the principle of inoculation, helping people recognize common mis- and disinformation formats or tactics, but it is acknowledged that this only works in certain circumstances and for certain people, and there is no universally effective solution.¹⁰⁵ Among these kinds of interventions are accuracy prompts, ‘prebunking’ and debunking, creating friction, encouraging lateral reading, providing media literacy tips, offering rebuttals to science denialism, self-reflection tools and learning about social norms as well as providing warning and fact-checking labels.¹⁰⁶ In

the case of lateral reading or thinking, this is often associated with education that aims to develop critical consciousness through education.¹⁰⁷ These interventions tend to be more audience-centered with less attention to the use of technical skills.¹⁰⁸

4.3 AI LITERACY

Encounters with information are increasingly shaped by the AI systems that generate information and personalize using algorithms that moderate and curate its flows. In this context, research focuses on the contribution that ‘AI literacy’ can make to MIL.¹⁰⁹ Several definitions of AI literacy can be found in the literature.¹¹⁰ One of the most cited is: ‘a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace’.¹¹¹

Whereas MIL typically refers to competencies required to use information and to communicate,¹¹² AI literacy definitions are likely to include the ability to comprehend the core principles and concepts of AI systems. Much effort has gone into defining what these competencies require in terms of specific skills. For example, training is expected to enable people to answer questions such as:

- What is AI? e.g., knowing differences between AI and other digital technologies.
- What can AI do? e.g., knowing what these differences mean for how AI can be used, its strengths and weaknesses.
- How does AI work? e.g., understanding the principal technical elements of AI.
- How should AI be used? e.g., understanding the ethical issues raised by use of AI.
- How do people perceive AI? e.g., understanding

¹⁰⁵ Kozyreva *et al.* (2024), funded in part by Humboldt Foundation, Volkswagen Foundation, European Union Horizon program, European Research Council (ERC), Australian Research Council (ARC) and Agence nationale de la recherche (ANR), France. See Section 4.2, Chapter 2 for a discussion of inoculation effects models. Fact-checking is discussed further in Section 2.1, Chapter 7.

¹⁰⁶ Kozyreva *et al.* (2024) funded as above.

¹⁰⁷ See Freire (1974); Wineburg & McGrew (2019), supported by the Robert R. McCormick Foundation and Spencer Foundation, US.

¹⁰⁸ For earlier literature reviews on media literacy, see Jeong *et al.* (2012), funded in part by the National Cancer Centre, South Korea; see also Livingstone (2008); on fact-checking, see Adjin-Tettey (2022), supported by the National Research Foundation of South Africa; on MIL, see also Frau-Meigs (2022). UNESCO has set global standards for MIL (2022a).

¹⁰⁹ Okunlaya *et al.* (2022). Some scholars are calling for ‘algorithmic literacy’, which has been in use in the literature since 1985, when it was defined as the ability of people without technical training to recognize when they interacting with a system driven by algorithms, to reason about what kinds of data might be collected, and to respond based on their decisions about how they want to interact with these systems; see Boots *et al.* (2024).

¹¹⁰ For a discussion of AI-powered interventions to counter mis- and disinformation, see Chapters 3 and 7.

¹¹¹ Long & Magerko (2020, p. 2).

¹¹² Henderson & Corry (2020).

common misconceptions about AI; making sense of AI; trustworthiness of AI.¹¹³

These competencies exemplify a broad view of AI literacy. When AI literacy is considered from the perspective of its relevance to MIL, it is suggested that a more specific set of skills or ‘micro competencies’ is important. These are divided into media (‘understanding the context of production’), documents (‘mastery of information search’) and data (‘oversight of algorithmic patterns’). In this context:

Media competencies include:

- Knowing the new context of news production and amplification via algorithms
- Being suspicious and aware of ‘weak signals’ of disinformation
- Fighting confirmation biases and other cognitive biases.

Document competencies include:

- Setting limits to tracking to reduce targeting (as fewer data are collected from devices)
- Browsing anonymously (e.g., use of virtual private networks, VPNs).

Data competencies include:

- Paying attention to platform adherence to data protection rules
- Mobilizing for more transparency and accountability about the impact of data use
- Signaling or reporting to platforms or web managers if data misuses are detected
- Commenting and/or rectifying ‘fake news’, whenever possible
- Alerting fact-checkers, journalists or the community of affinity.¹¹⁴

These competencies are intended to encourage a more proactive and potentially empowering approach to combating mis- and disinformation. In other words, they are not just about coping with mis- and disinformation, but rather boosting competencies for checking for possible exposure

to mis- and disinformation (e.g., be suspicious and aware of ‘weak signals’) and competencies designed to reduce the risk that exposure will occur in the first place (e.g., set limits to tracking so as to reduce targeting, such as ensuring fewer data are collected from devices, using VPNs [Virtual Private Network] etc. or contributing to data governance).

AI literacy as a tool for empowerment and resistance to the increasing control that AI systems exercise over people’s activities is a theme that is commanding some attention.¹¹⁵ Practical solutions are essential to address people’s lack of awareness about how algorithms make decisions that inhibit their agency. Bottom-up solutions are needed to respond to the ‘challenge of algorithm opacity by looking at the end-user (not the producer) and empowering citizens to analyse algorithms critically and creatively, in the hope of bringing insights in their own information consumption’.¹¹⁶

Empowerment and resistance depend as much – if not more – on paying attention to the inputs of AI systems as to their outputs. Hence, the term ‘data literacy’ is used to refer to the key set of competencies that people need to exercise control over their personal data, including what they allow to be collected and with whom it is shared. This is reflected in UNESCO’s definition of AI literacy as a combination of algorithmic literacy and data literacy:

‘[AI literacy] comprises both data literacy, or the ability to understand how AI collects, cleans, manipulates, and analyses data; and algorithm literacy, or the ability to understand how AI algorithms find patterns and connections in the data, which might be used for human-machine interactions’.¹¹⁷

Government bodies, international and civil society organizations are making concerted efforts to promote AI literacy. The governments of China, Germany, India, the United Arab Emirates (UAE) and the United States have put initiatives in place: for

¹¹³ Long & Magerko (2020).

¹¹⁴ Frau-Meigs (2024a), funded by the European Commission.

¹¹⁵ Stamboliev (2023), supported by the Vienna Science and Technology Fund (WWTF, Wiener Wissenschafts-Forschungs- und Technologiefonds), Austria.

¹¹⁶ Frau-Meigs (2024a, p. 512), supported by the European Commission.

¹¹⁷ UNESCO (2022b; emphasis added).

example, the Ministry of Education of the People's Republic of China's 2019 initiative and the Federal Ministry of Education and Research Germany initiative in 2021, and work which is ongoing in India. The UAE has had a structured program in place since 2018; and the 2020 National Artificial Intelligence Initiative in the United States.¹¹⁸ There are also initiatives throughout South-East Asia and in the European Union.

The focus of these AI literacy programs tends to be on their integration within existing primary, secondary and tertiary education. It is vital, however, that the needs of older people are provided for, with programs that take account of evidence that they are generally less digitally literate and so start from a lower base.¹¹⁹ Equally important is that AI literacy programs keep pace with the rapid advances in AI technologies, some of which are already being integrated into tools that billions of people use many times a day. For example, major internet search platforms (e.g., Google, Bing) are using GenAI to provide summarized responses to searches. The appeal for users is no longer having to examine the list of links that a traditional internet search produces, and to assess the quality and relevance of the results with respect to the answers they are looking for. GenAI summaries of search results will do that for them. The risk is that, as products of GenAI, these summaries may provide an inaccurate or even false (hallucinated) representation of the results.¹²⁰

One view is that GenAI tools such as ChatGPT and DALL-E are part of an incremental process of technological innovation. Sometimes it is argued that efforts to regulate these tools to combat mis- and disinformation should not be introduced if there is a risk that they will slow the rate of technological change.¹²¹ In this context, AI literacy training may offer an attractive option to those who resist regulatory interventions. This is especially so if it

is based on shared competency development and assessment criteria, and is effective in producing innovations in GenAI tools that operate in ways that are consistent with democratic processes.¹²²

4.4 EFFECTIVENESS OF MIL AND AI LITERACY INITIATIVES

Whatever content moderation practices are adopted by platforms, and given some of the evidence in some countries of a preference expressed by social media users for personal content moderation over platform moderation, there are likely to be continuing efforts to ensure that the public is in possession of MIL and AI literacy skills that could enable them to detect mis- and disinformation, interpret it, and exercise agency in their choices about how they respond.

Studies of the effectiveness of MIL interventions assessed in this report overrepresent the Global North. Where measures to combat mis- and disinformation have been tested globally, for example, debunking, accuracy prompts and media literacy tips, they are found to be sensitive to cultural contexts. The long-term effects of these measures have been tested to a limited extent, with some evidence suggesting that their effectiveness decreases with time. Comparative research in this area is limited by significant variability in methodologies (e.g., test stimuli using news headlines, real-world claim or websites) and different ways of measuring outcomes (e.g., belief or credibility ratings, behavioral measures).¹²³ Here are some examples:¹²⁴

- One analysis of media literacy interventions identified positive outcomes in relation to knowledge, criticism, influence, realism, beliefs, attitudes, norms, self-efficacy and behaviors, concluding that interventions were generally effective. The caveats were that effects might

¹¹⁸ Laupichler *et al.* (2022); Stanly (2024); UAE (2018); US Government (2020).

¹¹⁹ Loos & Ivan (2023); Moore & Hancock (2022).

¹²⁰ See Maynez *et al.* (2020).

¹²¹ Ross Arguedas & Simon (2023).

¹²² Ng *et al.* (2021); see also Chapter 5 on AI literacy.

¹²³ Kozyreva *et al.* (2024), based on 81 papers. Funded in part by Humboldt Foundation, Volkswagen Foundation, European Union Horizon program, European Research Council (ERC), Australian Research Council (ARC) and Agence nationale de la recherche (ANR), France.

¹²⁴ There are many reports and academic articles in the literature for countries which we do not have the resources to include in this report.

be greater for outcomes related to knowledge and realism than for attitude and behavior outcomes. Interventions were more likely to be successful if reinforced over multiple sessions.¹²⁵

- Civic education plays a role in MIL by increasing political efficacy, also a predictor of interest and trust in news. Students participating in a civic education program in the *United States* during 2003 and 2004 reported greater self-efficacy, an effect carrying over to increased political attentiveness and knowledge of candidate positions, with political attentiveness increasing knowledge and voting.¹²⁶
- In emerging democracies, civic education has been found to have favorable effects on levels of political information and participation, including a reduction in authoritarian nostalgia and an increase in desired political behavior.¹²⁷
- Experimental research in post-Soviet countries (*Belarus, Moldova, Russia and Ukraine*) found that civic education programs led to an increase in young people's support of democratic institutions, democratic attitudes and perceived political efficacy, albeit with small effects.¹²⁸
- An online focus group-based study in *Spain*, where participants were chosen proportionally to reflect the population, found that three main factors influenced the reported credibility of mis- and disinformation: channel (how an individual knows or discovers the content); source (provenance); and content (including topic and how it is conveyed). Women were found to be more vulnerable to mis- and disinformation than men. Older, better-educated, better-off participants, participants spending less time on the internet and those identifying as left-wing were less vulnerable

(only the effects of gender and age were statistically significant).¹²⁹

A systematic review of research on the impact of media literacy on young people's lives and well-being found some evidence that better internet skills are associated with thinking more about the credibility of online information.¹³⁰ Although evidence on the effectiveness of MIL strategies is absent in many countries, there does appear to be a common theme: this is the need to include both teachers and learners in acquiring values and practices consistent with respect for others and the need for 'whole-school' approaches to ensure that curricula revisions are culturally appropriate.

Literacy strategies with varying results.

In *Cote D'Ivoire*, a study demonstrated that Facebook's most widely shared posts were assassination rumors, vaccine skepticism, xenophobic hate speech and doxing of political opponents. Literacy training was introduced, but assessment showed that the intervention did not change how people consumed or shared information. In this case, the importance to online users of affirming group identities was found to take precedence over whether information was misleading.

In *Mali, Burkina Faso and Niger*, efforts to tackle online gender violence against women and girls illustrated the importance of collaborations between organizations that specialize in information verification and those involved in humanitarian and/or development initiatives to combat the flow of sexist information.

¹²⁵ Jeong *et al.* (2012), funded in part by the National Cancer Center, South Korea.

¹²⁶ Pasek *et al.* (2008).

¹²⁷ See Finkel (2014), sponsored by the US Agency for International Development (USAID); evidence from four evaluations sponsored by the USAID conducted since the late 1990s; Finkel *et al.* (2024), funded by the Middle East Partnership Initiative (MEPI) of the US State Department and partly by the European Research Council (ERC).

¹²⁸ Pospieszna *et al.* (2023), supported by the Polish National Science Centre. Sample participants were supportive of democracy before the treatment, although they were from non-democratic countries.

¹²⁹ Martínez-Costa *et al.* (2023), with 23 participants selected proportionally by gender, age (14–55+), income level, education level, level of internet use, political beliefs and geographical location, funded by the BBVA Foundation and European Commission.

¹³⁰ Livingstone *et al.* (2023a).

In *Cameroon*, a strategy to promote media and information literacy through awareness-raising for women emphasized the need to recognize power structures if inclusivity was to be achieved. Repressing harmful content was found to be difficult to achieve.¹³¹

Evidence of the impact of AI literacy training and levels on people's interaction with and consumption of online news is limited. There is evidence of growing awareness among digital system users of a wide range of potentially detrimental impacts of uncontrolled data collection on individual privacy that, in some contexts, may be linked to a rise in AI and data literacy. Concerns about data surveillance – 'the monitoring of citizens on the basis of their on-line data' – are manifest in growing resistance to the data collection policies of digital system providers.¹³² In the work environment, this may take the form 'gaming the system' by entering fake data inputs, through to opposition to the adoption of algorithmic management practices. Beyond work, resistance ranges from citizens paying more attention to managing their privacy settings on digital platforms, changing the ways they use these platforms, or opting out of purchasing certain digital consumer products.

Some researchers cast doubt on the efficacy of these resistance practices: 'it is questionable whether an average user can actually accomplish the task of understanding and recognizing all risks and challenges related to privacy in an increasingly complex and ever-changing media environment'.¹³³

It is clear in much of the research literature on media, information, data, digital or AI literacy that 'critical' literacy is essential. In addition to the practical challenges of generating and circulating information, producing less biased data sets and understanding how algorithms work, it is important for individuals to be able to make sense of information. If the critical literacies of populations

are developed, this is likely to influence decisions about whether there is 'equality and/or symmetry between human and non-human actors, and ... [the] conceptualization, development and understanding of new forms of intelligence we would like to live with in the future'.¹³⁴ Those with critical literacies can encourage imaginative approaches to the design and use of AI systems, addressing ethical issues and recognizing when AI tools, data collection and processing are helpful and when they are not.¹³⁵

5 Chapter Summary

With the growing concern about harmful consequences of corporate datafication strategies and increased flows of mis- and disinformation, there are numerous efforts to measure its scale and to assess how people in different parts of the world engage with and respond to it. Measurement is difficult in the absence of access to the platforms' data, which also means the severity of impacts on individuals and society is also extremely difficult to measure. Further challenges are due to the fact that mis- and disinformation are produced and circulated outside social media. This complicates the identification of impacts that can be misleadingly attributed to the role of social media and the digital platforms' algorithms.

Concern is also growing about the contribution of mis- and disinformation to infringements of rights, especially of those of children. The owners of the largest platforms, AI systems developer companies and the advertising industry, are promoting the virtues of online engagement for the young and the old. These companies claim to be acting responsibly with regard to the rights of their users. However, their reported failures to do so means that information integrity and the sustainability of healthy information ecosystems are increasingly high on policy agendas at the highest levels.

¹³¹ Birwe (2024); NDI (2023); Zibi Fama (2024).

¹³² Resistance strategies are examined in detail in Chapter 8.

¹³³ Masur *et al.* (2021, p. 10).

¹³⁴ Jandrić (2019, p. 35); see also Ritzer *et al.* (2024).

¹³⁵ Deuze & Beckett (2022).

This chapter has addressed evidence on the scale and severity of mis- and disinformation and how well the public (and the policy maker community) understands the way mis- and disinformation can influence what people believe and how they behave online and offline. While countries are struggling to impose rules on big tech companies that aim to alter how algorithms operate to reduce the excesses of mis- and disinformation, improving literacy is seen as an attractive option for policy makers and for the big tech companies. Literacy-improving measures are garnering renewed attention now that they have been coupled with measures to improve AI literacy. However, MIL and AI literacy policies must be accompanied by both state-led and individual or community-led responses to information problems created by corporate datafication practices.

The synthesis of research in this chapter shows that:

- Gauging the scale and severity of mis- and disinformation is difficult due to challenges in collecting and analyzing data that reflects people's online experiences. Privacy protection and ethical issues and big tech company restrictions on access to data increase the challenges of measurement. The experience of mis- and disinformation is influenced by conditions in people's offline lives in ways that are neglected in studies that focus primarily on information itself.
- Evidence on the scale and severity of harms associated with mis- and disinformation comes mostly from quantitative surveys and experimental research. Large-scale studies are limited to a few platforms, and are largely centered on the United States.
- Evidence on children's susceptibility to mis- and disinformation and its impact on their rights and well-being is less well developed than research on adults.
- Research on public awareness of the role of AI systems in generating and circulating mis- and disinformation reaches different conclusions depending on the criteria used and on the context. Studies reveal that people have varying levels of confidence (whether justified or not) in their capacities to identify AI generated mis- and disinformation, and evidence at the population level is relatively weak.
- There is considerable variation in self-reported understanding of AI systems and algorithms, their use in the production of news media and how these affect people's lives.
- Research on people's acceptance of interventions by governments or companies to tackle mis- and disinformation varies by country and context. There is uncertainty about who is responsible for rights protections, that is, the state, big tech companies or individuals themselves.
- Policy makers need to develop an improved understanding of AI systems and digital technologies generally. Issues are often addressed in institutional silos. There is little systematic research on what policy makers understand about the multiple factors that contribute to a mis- and disinformation 'crisis'. Policy makers, especially in Global Majority World regions, are said to lack resources to address harms to individuals and democratic processes. Lobbying by big tech companies can lead to ad hoc policy and have a chilling effect on freedom of expression.
- MIL and AI Literacy initiatives should focus on more than technical skills, and literacy initiatives should not be seen as a sufficient response to mis- and disinformation.
- Adults and children who have engaged with critical literacy training are more likely to be able to differentiate between legitimate and other sources of information, and to participate in making choices about the design and use of digital systems, including AI systems.
- MIL interventions for countering mis- and disinformation can lead to improvements in how people engage with online information.

Encouraging positive outcomes have been reported. However, literacy training alone should not be regarded as a complete answer to mis- and disinformation problems.

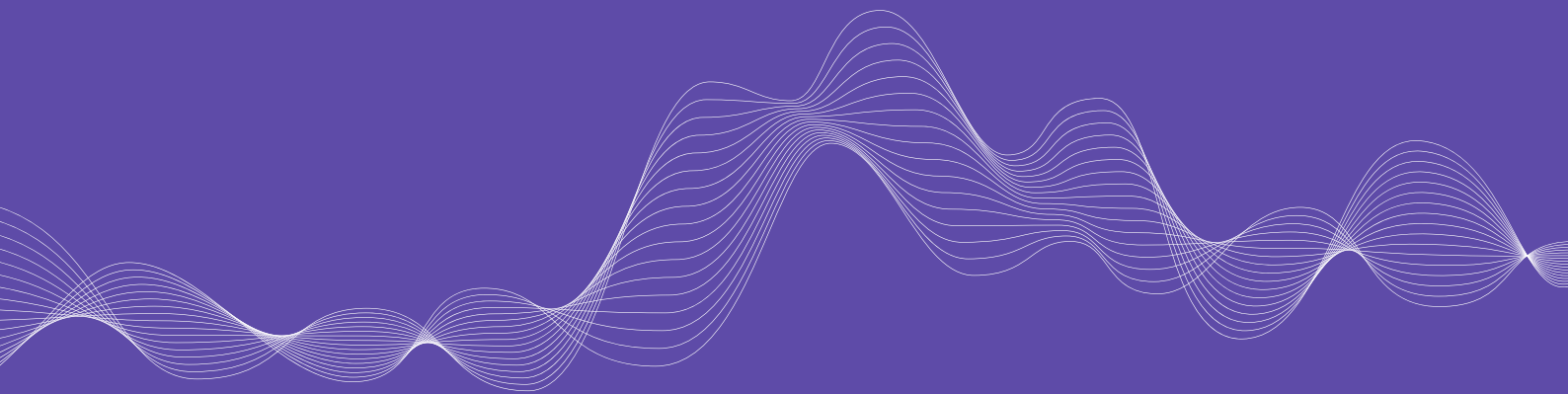
- AI literacy definitions are being developed that combine algorithmic literacy with data literacy. Enhancing AI literacy is crucial at all stages of AI systems development, and deployment and lifelong learning programs are essential.

Research is needed:

- To provide improved measures of the scale and perceptions of the severity of mis- and disinformation in countries around the world and over time while ensuring research is conducted legally and ethically.
- To understand the interactions that influence the severity of harms associated with mis- and disinformation and to extend research to a larger number of platforms – both large and small – systematically and outside the United States and Europe.
- To understand the conditions that lead to differences in people’s reported concerns about their online safety and their confidence in identifying mis- and disinformation.
- To investigate ways of improving communication to the public about who is responsible for protecting their rights.
- To evaluate how policy makers can acquire a better understanding of the societal conditions that give rise to illegal and harmful mis- and disinformation.
- To develop standardized MIL and AI literacy definitions and cross-country comparative conceptual frameworks and methodologies to support research on the effectiveness of initiatives that respect different cultures and values.
- To understand how critical literacy skills training can be taught effectively to children and adults.

CHAPTER 6

GOVERNING INFORMATION ECOSYSTEMS: LEGISLATION AND REGULATION



The map is generated by the OID on the basis of the chapter's text using GarganText – developed by the CNRS Institute of Complex Systems. Starting from a co-occurrence matrix generated from chapter's text, GarganText forms a network where words are connected if they are likely to occur together. Clustering is conducted based on the Louvain community detection method, and the visualization is generated using the Force Atlas 2 algorithm.

i

This chapter provides an account of selected legislative and regulatory tools that are available to governments to mitigate the harms of mis- and disinformation and to govern the way mainly big tech companies operate.¹

The research synthesis focuses on:

- **What types of governance approaches are available?** This briefly discusses voluntary governance that relies on corporate self-regulation and anticipatory co-regulatory and direct state regulatory approaches as well as remedial approaches such as competition/anti-trust measures.
- **What approaches to information ecosystems governance are being promoted at the global level?** This highlights principles that are being established for governing information ecosystems and the emphasis given to human rights protections.
- **What are some of the legislative, regulatory and judicial approaches to governing information ecosystems?** This explains governance approaches applied at regional or national levels. Anticipatory and remedial approaches are discussed: network neutrality policies aiming to open the digital infrastructure; privacy and data protection measures; digital platform regulation; and the regulation of AI systems and news media.

This chapter emphasizes normative goals and rules embodied in selected governance approaches, providing an insight into tensions between these goals and rules and their implementation in view of the interests of different actors.

Chapter 7 examines how governance practices are being deployed to combat mis- and disinformation to strengthen the health of information ecosystems. Chapter 8 critically examines alternative data governance practices aimed at resisting injustices, biases and the harms of big tech-enabled datafication practices.

¹ For background reading, see Flew (2021). A comprehensive analysis of research in this area is beyond the scope of this report. See Appendix: Methodology for details of literature review process.

1 Introduction

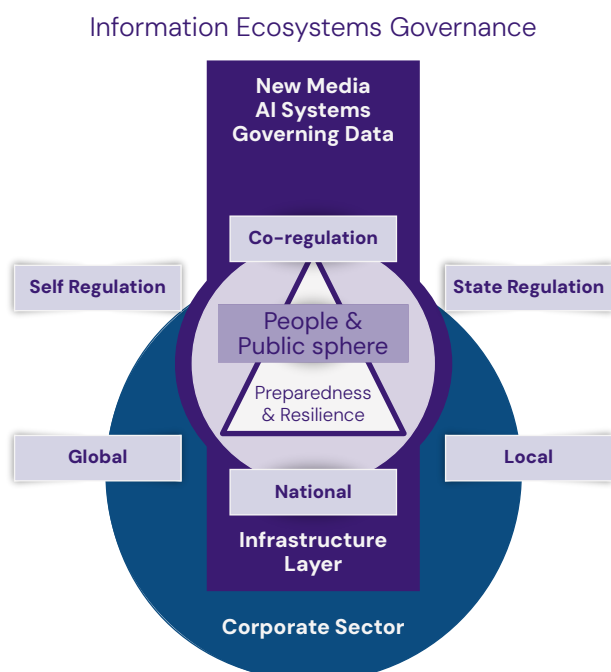
Governance of information ecosystems is concerned with both the role of governments to cultivate ‘systems, institutions, and norms that enable quality and useful information to flourish’,² and also the corporate actors that supply the digital technologies that are implicated in the spread of mis- and disinformation. ‘Governance’ is understood here broadly to encompass patterns of rules that underpin social orders.³

The focus in this chapter is on interdependent systems and the multiple interactions of information flows, technology and communication infrastructures, norms and practices of public and private institutions. Governance approaches are complicated by the fact that information ecosystems are composed of layers, each with its own conditions and actors, both public and private.⁴ These layers – infrastructure and applications – in this report, support news media and AI systems and the way data is governed (see Figure 6.1).

Figure 6.1 locates the preparedness and resilience of people and their communities at the center of information ecosystems, indicating that governance arrangements – principles and institutions – are put in place globally, regionally or nationally through state regulations, co-regulatory measures involving states and the private sector, and the self-regulatory initiatives of companies that supply the information ecosystems infrastructure layer and support the applications such as hosting news media content. In this chapter we look principally at the governance of corporate actors (recognizing that in some countries state ownership plays a crucial role, and that local, municipal and community civil society actors are also taking initiatives to govern information ecosystems).

2 Types of Governance Approaches

Figure 6.1
Simplified view of the governance of information ecosystems



Source: Authors of this report

Voluntary governance measures rely on self-regulation by the tech company owners whose platforms host and circulate content, and by the companies that invest in and operate the underlying infrastructure, including the internet.⁵ In this case private actors are expected to commit voluntarily to more stringent standards of practice, consistent with norms and values agreed internationally, regionally and nationally. Power asymmetries due to the monopolistic practices of many of these companies result in clashes between business interests and the public interest. Self-regulation may be intended, for instance, to protect the integrity of elections and the health of democracies, but it is limited due to its voluntary nature and the

² Radsch (2023a).

³ ‘Governance’ may refer to formal rules initiated by states, corporate self-regulation or co-regulation as well as informal rules and norms put in place individually or collectively. See Puppis *et al.* (2024) on multiple ways in which the term ‘governance’ is used and criticized in the literature.

⁴ The layers of information ecosystems can be depicted in many different ways. See van Dijck (2020), for example, a depiction of the components of digital platforms as a tree structure with roots and branches organized differently in the in the European Union and the United States.

⁵ Kokshagina *et al.* (2023), supported by the European Commission.

possible interference with business interests (i.e., to maximize and monetize user online engagement) or by indifference to the public's interest (i.e., in the protection of human rights or the maintenance of healthy information ecosystems).⁶

Most digital platforms employ some form of content governance. They claim that their practices embody fundamental human rights protections, including freedom of expression and privacy protection. Typically, they have no dedicated policy specifically regarding mis- and disinformation, yet it is these big tech companies that decide how mis- and disinformation are addressed.⁷ Recent efforts to achieve international consensus on what is expected of corporate self-regulatory governance are discussed in Section 3 of this chapter. Governance approaches applied at regional and national levels to network infrastructure, and for privacy and data protection, digital platform and AI systems governance and news media regulation are discussed in Section 4.

States are primarily responsible for acting to protect human rights and fundamental freedoms, including in the digital environment. They have a *negative obligation* not to violate rights – including the rights of digital companies. They also have a *positive obligation* to protect human rights and implement them in practice. Every intervention is therefore a balancing act that must be assessed in each specific context.⁸ While governance in the form of state and co-regulation can address certain illegal acts, most mis- and disinformation cannot simply be banned because much of it is not illegal per se.⁹

When voluntary self-regulation does not address concerns about the way the corporate sector is operating, *anticipatory governance* measures are used. These introduce legislative obligations that must be adhered to by the companies

developing and operating components of information ecosystems. They may take the form of *co-regulation* (state and corporate), which is becoming common as concerns about the power of digital platforms and other data intermediaries are growing, and voluntary mis- and disinformation countermeasures are deemed insufficient to mitigate harms.¹⁰ This approach aims to correct the power asymmetry between the owners of digital platforms and other actors in the ecosystem. Typically, co-regulation takes the form of regulations applied to dominant firms to establish norms and rules for their behavior.¹¹ It is regarded as a potentially balanced option – between the interests of the public and the interests of companies in succeeding in the commercial market. When the states legislate to set up co-regulatory arrangements, this can also involve participation by the private sector and some form of civil society representation in decision-making processes.¹²

Anticipatory governance in this form is seen in some regions as more flexible and inclusive than direct state regulation. This is because co-regulation is said to leave less room for abuse and discretionary measures on the part of the state.¹³ *State regulation* involves governments enacting legislation that grants them the authority to decide how information ecosystems should be structured and managed, which can result in rights-infringing measures and partisanship.

In addition, a remedial form of governance led by the state in the form of competition/anti-trust measures is becoming more common. This is premised on the view that competing infrastructure and service providers is consistent with the public interest. The state can also undertake other remedial actions, such as legislating changes in ownership arrangements for data or news organizations.

⁶ De Blasio & Selva (2021).

⁷ See Chapter 2 for news media and Chapter 3 for AI systems.

⁸ Tenove (2020), funded by the Social Sciences and Humanities Research Council (SSHRC) of Canada.

⁹ For an introduction to various forms of regulation, see Brown & Marsden (2023).

¹⁰ Self-regulation was still the preferred approach in 2021 in the Czech Republic, which addresses problems of mis- and disinformation through intelligence strategies; see De Blasio & Selva (2021).

¹¹ Pickard (2020b).

¹² See De Blasio & Selva (2021) on state/industry co-regulatory bodies, technical measures and codes of conduct.

¹³ Dittrich (2019); Durach et al. (2020).

3 Global Governance of Information Ecosystems

Concerns about an ‘information crisis’, political polarization, harms to individuals and groups and the destabilization of democracies in the wake of datafication for profit have led to global initiatives to address these concerns. Concluding that declining trust in major institutions globally is partly due to failure to provide reliable information, in 2023 the United Nations proposed a voluntary code of conduct relating to the integrity of information ecosystems (see Table 6.1):

All stakeholders should refrain from using, supporting or amplifying disinformation and hate speech for any purpose, including to pursue political, military or other strategic goals, incite violence, undermine democratic processes or target civilian populations, vulnerable groups, communities or individuals.¹⁴

UNESCO’s governance initiative takes the form of *Guidelines for the Governance of Digital Platforms*. It establishes voluntary principles and guidelines for duties, responsibilities and roles for stakeholders, with the aim of safeguarding freedom of expression, access to information and other basic human rights (see Table 6.1). It also sets out guidance for policy makers for addressing hate speech through education.¹⁵

Table 6.1
Governing information ecosystems

Principles for information integrity: United Nations	Principles for governing digital platforms: UNESCO
<ul style="list-style-type: none"> • Commitment to information integrity • Respect for human rights • Support for independent media • Increased transparency • User empowerment • Strengthened research and data access • Scaled-up responses • Stronger disincentives • Enhanced trust and safety 	<ul style="list-style-type: none"> • Platform owners to conduct human rights due diligence • Platform owners should adhere to international human rights standards, including in platform design, content moderation and content curation • Platform operations are transparent • Platform companies make available information accessible • Platform owners are accountable to relevant stakeholders

Source: UN (2023a) and UNESCO (2023b)

Other intergovernmental organizations have stepped up efforts to mitigate threats associated with mis- and disinformation. For example, the OECD observed in 2024 that:

What makes content-specific regulatory responses particularly complex is not only that defining what content may be restricted without infringing upon freedom of expression is difficult, but also that illiberal regimes can co-opt laws to combat disinformation developed in countries with effective checks and balances to legitimise their own antidemocratic practices.¹⁶

The OECD is working towards a framework that would help to enhance the transparency, accountability and plurality of information sources; foster societal resilience; upgrade governance measures; and encourage institutional arrangements that uphold the integrity of the information space. Bilateral initiatives aim to form coalitions among like-minded countries, for example a United States-led effort aimed at protecting democracies from the disinformation campaigns of foreign governments.¹⁷

¹⁴ UN (2023a).

¹⁵ UNESCO (2023b). The Guidelines were produced through a multistakeholder consultation, gathering more than 10,000 comments from 134 countries; see also UNESCO (2023a), a multi-stakeholder consultation that received 10,000 comments from 123 countries.

¹⁶ OECD (2024).

¹⁷ Wintour (2024), announced in April 2024, and signed by Canada, the United Kingdom and the United States.

The framing of the impact of digital platforms and their technologies, including AI systems, on information ecosystems in governance contexts depends on ‘political decisions about normative issues’,¹⁸ reflecting the interconnected nature of digital technologies and societal norms. Developing rules for news media, digital platforms and AI is a key means of exercising normative influence over global regulation, and all these initiatives are framed by international human rights rules, even if the best means of institutionalizing these rules is contested.¹⁹ At the global level, these contests among stakeholders – public and private, individual and collective – played out in deliberations that led in 2024 to the United Nations’ *Pact for the Future*, setting out ‘guiding principles’, which, among others, embrace ‘full respect for international law’, ‘the pursuit and enjoyment of human rights and fundamental freedoms for all’ and ‘the responsible and ethical use of science, technology of innovation, guided by the principles of equity and solidarity’.²⁰ Annex I is a *Global Digital Compact* that sets out five objectives:

1. Close all digital divides and accelerate progress across the Sustainable Development Goals;
2. Expand inclusion in and benefits from the digital economy for all;
3. Foster an inclusive, open, safe and secure digital space that respects, protects and promote human rights;
4. Advance responsible, equitable and interoperable data governance approaches;
5. Enhance international governance of artificial intelligence for the benefit of humanity.²¹

The actions include addressing connectivity and digital divides, addressing digital literacy, skills and capacities, promoting digital public goods and digital public infrastructure, expanding inclusion in the digital economy and promoting the ‘free flow of information and ideas’, calling on digital technology companies to respect international human rights and principles.²²

Regarding digital trust and safety, the Compact states that:

We must urgently counter and address all forms of violence, including sexual and gender-based violence, which occurs through or is amplified by the use of technology, all forms of hate speech and discrimination, misinformation and disinformation, cyberbullying and child sexual exploitation and abuse. We will establish and maintain robust risk mitigation and redress measures that also protect privacy and freedom of expression.²³

The Compact explicitly refers to information integrity:

We will work together to promote information integrity, tolerance and respect in the digital space, as well as to protect the integrity of democratic processes. We will strengthen international cooperation to address the challenge of misinformation and disinformation and hate speech online and mitigate the risks of information manipulation in a manner consistent with international law.²⁴

In this context, specific commitments to be achieved by 2030 include: ‘digital media and information literacy curricula’, promoting ‘diverse and resilient information ecosystems’, including the strengthening of independent and public media as well as supporting journalists and media workers, and providing, promoting and facilitating ‘access to and dissemination of independent, fact-based, timely, targeted, clear, accessible, multilingual and science-based information’, along with other commitments.²⁵ Other issues addressed under other objectives include data privacy and security, standards, data flows and AI.

¹⁸ Erman & Furendal (2022, p. 267), supported by the Marianne and Marcus Wallenberg Foundation and Swedish Research Council (Vetenskapsrådet).

¹⁹ Roberts *et al.* (2024).

²⁰ UN (2024b, pp. 57–58). For the full list of guiding principles and commitments, see pp. 58–60.

²¹ UN (2024b, pp. 40–41).

²² UN (2024b, pp. 41–56).

²³ UN (2024b, p. 48).

²⁴ UN (2024b, p. 49).

²⁵ UN (2024b, p. 49).

These statements of commitments necessarily are voluntary and have less traction than the governance rules that are introduced at national level by states or by regions and through the self-regulatory initiatives of globally operating companies.

4 Governance Approaches Applied at Regional and National Levels

This section explains governance approaches that are developed and applied at regional or national levels, although they are informed by commitments to voluntary principles that are agreed at the global level. We start with a selection of both anticipatory and remedial approaches to governance in areas that are expected to impact on the health of information ecosystems, beginning with network neutrality measures designed to secure an open internet (Section 4.1). We then review privacy and data protection measures (Section 4.2), digital platform regulation (Section 4.3), AI systems regulation (Section 4.4) and finally, approaches to news media regulation (Section 4.5).

4.1 GOVERNING NETWORK INFRASTRUCTURE

Internet connectivity and access are central to how people experience information ecosystems. While we cannot address all the features of governance in this area, network neutrality policies and regulations concerning what is known as ‘zero rating’ are central to how those who do have connections and affordable access experience information ecosystems. This form of anticipatory governance typically involves legislation and co-regulation, but it can also involve the state acting authoritatively under legislation that permits it

to set rules for internet access and use. Whether network neutrality rules are adhered to conditions whether and how people can generate and amplify the circulation of all kinds of information, including mis- and disinformation, and what information they encounter online.

Network neutrality is the principle that internet service providers (ISPs) should treat all data (information) that flows through their networks without discrimination. This open internet principle is controversial because it impacts on the equality of access to data and online information.²⁶ The principle emerged in the Global North, and it intersects with zero-rating practices that are now common in many countries in the Global Majority World — ISPs offer access to certain services, and data usage does not count against a cap on the data used to access those services. This means that owing to a desire to minimize costs, users may restrict their access to information to a limited number of platforms.

Approaches to network neutrality and zero rating in India. Network neutrality and zero rating became critical policy issues in *India*, attracting intense scrutiny from online content firms and ISPs. The Telecom Regulatory Authority of *India* (TRAI) consulted on network neutrality in 2016 and 2017, after a #SaveTheInternet campaign by activists against Facebook’s Internet.org. Facebook aimed to provide low-cost and subsidized access to a few selected services to lower-income countries in Asia and Africa. This zero-rating service raised concerns about fairness and competition, because it would give preferential treatment to certain services over others.

Network neutrality debates focused on traffic management practices, that is, the prioritization of certain types of internet traffic over others, potentially disadvantaging

²⁶ For a discussion, see Baranes (2014); Bauer & Knieps (2018); Economides & Hermalin (2012); Hildebrandt & Wiewiorra (2024); Jordan (2017); Marsden (2016); Marsden & Brown (2023); Menon (2021); Pickard & Berman (2019); Winseck & Pooley (2017); Wu (2003); Yoo (2024). For a literature review, see Lee & Shin (2016).

some users or competing services. The conflict is between those calling for an open internet, where all data is treated equally, and the ISPs that claim they need to manage network traffic under conditions of congestion and to block illegal content. TRAI banned discriminatory tariffs in 2016, effectively prohibiting zero-rating service offerings like Facebook's Free Basics.²⁷

In India's diverse socio-economic context, where internet access is a critical developmental tool, its policy of ensuring equal access by banning a two-tiered internet illustrates how a balance may need to be struck between equity and innovation in service provision. The Cellular Operators Association of India (COAI) suggests that the policy is limiting the introduction of lower-cost access that might help to bridge the digital divide.²⁸

Network neutrality and zero-rating issues are widely discussed in *South Africa*, *South Korea* and *Latin America* in relation to the public value of the internet, and where digital activism aims to resist Facebook's Free Basics service.²⁹ In jurisdictions allowing zero rating, regulatory bodies, such as the Independent Communications Authority of South Africa (ICASA), provide guidelines to try to align these practices with public interest goals, for example to enhance educational and public health access to information or to prevent anti-competitive behavior.

In the *United States*, net neutrality policy is in regulatory flux. The Federal Communications Commission established strong net neutrality rules in 2015 so that ISPs could not discriminate between preferred online service providers. A Pew Research Center survey found that when network neutrality

rules were in place, a majority of Americans reported that they either understood or supported the policy even though its enforcement was inconsistent and impacted on the quality of service they experienced online.³⁰ Network neutrality was repealed in the United States in 2017, reinstated in 2024 and then blocked by the federal court. There is some evidence that without net neutrality rules, there has been an increase in ISP data throttling and prioritization, which can be argued to disadvantage smaller content providers and reduce consumer choice.

The *European Union's* Open Internet Access Regulation provides rules on net neutrality across member states, with the Body of European Regulators for Electronic Communications (BEREC) setting guidelines requiring that ISPs do not favor specific service providers. These indicate that zero-rating practices must not undermine net neutrality.³¹ Here, too, the policy is controversial; although there is a high compliance rate, critics argue that the policy allows for subtle traffic prioritization that is discriminatory.

Techniques for closing off internet access, whether via zero rating, throttling traffic or other means of fragmenting the internet, are also used widely.³² Strong measures include internet shutdowns and social media blocking during elections, with political unrest and protests occurring in countries as diverse as *Belarus*, *Iran*, *Myanmar*, *Turkey*, *Vietnam* and *Zimbabwe*.³³ Interference by authoritarian regimes includes restrictions on access to information, such as *China's* Great Firewall, heavily regulated 'national internets' and using intrusive content governance measures to favor or censor political speech or for surveillance,³⁴ such as, for example, *Iran's* initiative to create a 'national information network', requiring websites and services to locate servers inside the country and increasing the cost of global internet traffic.³⁵

²⁷ Eisenach (2015); Mukerjee (2016); Prasad (2018). For a more extensive treatment of zero rating, to give a sense of its scope and the debate around whether it addresses exclusion problems or unjustly reduces access to information, see Gerpott (2018); Hoskins (2024); Jauniaux & Lebourges (2019); Krämer & Peitz (2018); Mattelart (2023).

²⁸ Menon (2021).

²⁹ Nothias (2020); Robb & Hawthorne (2019); Shahin (2019); Shin & Lee (2017).

³⁰ Greenstein *et al.* (2016); Program for Public Consultation (2022); Vogels & Anderson (2019).

³¹ BEREC (2024a, b).

³² Boas (2006); Howard & Hussain (2013); Kalathil & Boas (2003); Shahbaz *et al.* (2022).

³³ Akser & Baybars (2023); Mare (2020); Sinpeng (2020); Ryng *et al.* (2022), supported in part by the Office of the United Nations High Commissioner for Human Rights (OHCHR).

³⁴ Keremoğlu & Weidmann (2020), funded by the German National Science Foundation (DFG, Deutsche Forschungsgemeinschaft).

³⁵ Motamedi (2024).

Debates in this area concerning the infrastructure layer highlight the need for adaptable and context-sensitive regulation so that the benefits of digital inclusion and access to diverse sources of information are balanced against risks of market distortions. One strategy is to require ISPs to disclose their data management practices and zero-rating agreements publicly, and to include civil society and industry in policy making. Decisions are increasingly influenced by efforts to achieve internet or digital sovereignty, which is understood differently depending on a country's political and economic context.³⁶

4.2 PRIVACY AND DATA PROTECTION GOVERNANCE

The collection and processing of sensitive personal and non-personal data on an industrial scale in the data economy by big tech companies means that governance rules are being updated to mitigate risks of privacy infringements and harms resulting from identity exposure. Specific rules apply for different types of data depending on their sensitivity and the risks associated with their misuse, with the aim of increasing transparency and accountability for data use.³⁷ In this area we find a mix of anticipatory and remedial governance measures.

In the United States the capacity of digital platforms to collect, process and make data generated online available to third parties without user consent is subject to privacy protection legislation at the federal level, with the Children's Online Privacy Protection Act (COPPA) of 1998 updated in 2013 to govern the collection of information about minors, addressing issues of parental consent, confidentiality and security, with safe harbor provisions and rules for data retention and deletion.³⁸ There is no single federal law to govern data privacy, but federal laws apply to data and telecommunications, health information, credit, financial and marketing information. There are

multiple state-level laws, including the California Consumer Privacy Act (CCPA). The Federal Trade Commission (FTC) functions as a regulator to constrain unfair or 'deceptive or unfair business practices and from unfair methods of competition', and takes action to enforce privacy laws.³⁹

- In the *United States*, the CCPA is seen as the most stringent privacy law. It requires businesses to disclose the categories and specific pieces of personal information they collect at or before the point of collection. It asserts the consumer's right to know about the personal information that is collected and when it is sold or shared with third parties. This model broadened the concept of data 'sale', potentially encompassing many types of data transactions not typically considered sales, and requiring businesses to reevaluate their data practices.⁴⁰
- In the *United States*, the Health Insurance Portability and Accountability Act (HIPAA) was introduced in 1996 with data privacy and security provisions for safeguarding medical information. The HIPAA, while comprehensive, does not fully address the complexities of new technologies and the digitization of health records. Mobile health applications and wearable technology are generating vast amounts of health-related data that can fall outside the scope of this legislation. This underscores the need for continuous enhancement of legal frameworks to keep pace with technological advances and societal changes. Data protection becomes even more complex when sensitive topics, such as access to abortion data, become an issue, and the handling of, for example, abortion data, comes under intense scrutiny. The reversal of *Roe v. Wade* has heightened concerns about the privacy and security of reproductive health data. The protection of such sensitive data is crucial in preventing data misuse and discrimination, and in ensuring that individuals' privacy rights are upheld.⁴¹

³⁶ Afina *et al.* (2024); Kokas (2022); Kumar & Thussu (2023); Stefanija & Pierson (2023).

³⁷ Kerber (2020).

³⁸ US Congress (2013).

³⁹ See FTC website <https://www.ftc.gov/about-ftc> and see Kira *et al.* (2021) for an overview, supported by the Global Challenges Research Fund (GCRF), UK and the Omidyar Network.

⁴⁰ US State of California (2018).

⁴¹ Dellinger & Pell (2024); Roth (2022); US Congress (1996a).

- The *European Union's* General Data Protection Regulation (GDPR) references fundamental rights enshrined in Articles 7 and 8 of its Fundamental Rights Charter, and is among the most comprehensive regulation worldwide.⁴² Implemented in May 2018, it sets out requirements for companies and organizations that collect, store and manage personal data. It applies uniformly across all sectors, but has specific provisions for some types of data-processing activities. For example, Article 9 imposes stricter conditions on the processing of special categories of personal data such as health information, biometric data and data revealing racial or ethnic origin.

A shift in tech company behavior was recorded when data collectors were required to disclose their data handling practices and undergo regular audits to ensure compliance with legal standards.⁴³

Compliance with the GDPR. Google has faced significant challenges in complying with the GDPR. The stringent demands for increased transparency and data handling accountability compelled the company to overhaul its privacy policies and practices. It revised its privacy policies to make them more understandable and accessible to users, simplifying the language and providing clearer explanations of what data is collected and how it is used. The policies now include detailed descriptions of privacy controls that users can access to manage their personal information, aiming to ensure users have a better understanding and greater control over their data. Google introduced more granular privacy controls in user account settings, allowing users to more easily review and modify privacy options. A proactive 'Privacy Checkup' tool was rolled out that guides users

through their privacy settings, aiming to help them make better informed decisions about their data.⁴⁴

Questions remain about whether the regulation goes far enough to ensure data privacy, and tech companies are frequently charged with data breaches and with unauthorized tracking of users online. For example, a case has been brought under European Union competition policy anti-trust rules law against Alphabet Inc. for Google's alleged tracking of users. In June 2024, action by the European Center for Digital Rights (NYOB), a privacy advocacy group, resulted in Google being scrutinized under European Union anti-trust law for unauthorized user tracking by its Chrome web browser.⁴⁵ Privacy controls on platforms such as Facebook have also been criticized for their use of so-called 'dark patterns' – 'tricks used in websites and apps that make you do things that you didn't mean to, like buying or signing up for something' all create opportunities for unauthorized data collection that are opaque to the user.⁴⁶ Alleged infringements of privacy can take a long time to resolve. For example, the *Irish* Data Protection Commission's 2019 investigation into whether Google uses sensitive personal data about race, health or political preferences to target ads stalled. The Irish Council for Civil Liberties then asked the Irish High Court to force an investigation, but this request was denied, although the Data Protection Commission did start an investigation in early 2024.⁴⁷

The complexity of data collection practices and the volume of data pose significant challenges for achieving true transparency, and data protection authorities are struggling to keep pace with multiple cases before the courts. The power asymmetry between large tech companies and their users can leave the latter unaware of the full extent

⁴² EC (2016b).

⁴³ Linden *et al.* (2020), supported in part by the National Science Foundation (NSF), US.

⁴⁴ See Houser & Voss (2018); Murtaza & Salman (2019); Waldman (2020).

⁴⁵ See Halpern *et al.* (2024).

⁴⁶ See Brignull (2023, p. np).

⁴⁷ See Murgia (2019); O'Faolain (2024).

and implications of data collection and data monetization; targeted advertising and the misuse of data for political gain mean that the GDPR is not a panacea for all data economy issues. Vigilance and a commitment to ethical data practice are essential to protect user privacy and maintain public trust.

While the GDPR has served as a template in several jurisdictions,⁴⁸ others have varied approaches, reflecting their unique socio-political and economic contexts:

- In *African countries* measures are being taken to introduce privacy protection and data protection legislation. A data protection law was put in place in *Cape Verde* in 2001. As of the end of 2023, 35 countries had enacted legislation, with three others pending, although a review indicates that in some cases countries introduce exemptions for national security reasons. While the GDPR in Europe also provides exemptions, the issue in African countries is the robustness of the institutional protection of human rights.⁴⁹
- *Brazil's* General Personal Data Protection Law (LGPD, Lei Geral de Proteção de Dados) mirrors the GDPR's comprehensive scope while incorporating elements tailored to the country's environment. The LGPD emphasizes principles of transparency, purpose limitation and data minimization, aiming to balance the protection of personal data with the facilitation of economic activities.⁵⁰
- *China's* approach to data protection, exemplified by the Cybersecurity Law and Personal Information Protection Law (PIPL), is tied to its broader strategy to balance the imperatives of economic growth and national security, reflecting its socio-political and economic landscape.⁵¹ It is argued in the critical literature

that the main emphasis of the country's data laws is on treating data as a 'new factor of production', which does not acknowledge people's epistemic rights, that is, their right to know.⁵²

- In *India*, a Draft Personal Data Protection Act (DPDPA) 2023 was proposed in 2018 after a landmark Supreme Court judgment – *Puttaswamy vs. Union of India* in 2017 – and passed in 2023. The DPDPA mirrors aspects of the GDPR, and is aimed at 'the processing of digital personal data in a manner that recognises both the right of individuals to protect their personal data and the need to process such personal data for lawful purposes'. In some sectors, such as financial technology (fintech), sectoral regulations apply. For example, the Reserve Bank of India plays an important role in regulating the financial services industry, establishing and enforcing self-regulatory guidelines, and penalizing and suspending bank licenses that do not comply with its data protection guidelines and 'know your customer' norms.⁵³
- *Japan* has updated its Act on the Protection of Personal Information (APPI) several times after its introduction in 2003, and achieved GDPR compliance a year before the European Union's legislation came into force. Further amendments have expanded the scope of individual rights, provided for stricter consent requirements, made data breach notifications mandatory, and limited the types of data that can be provided to third parties.⁵⁴

These rules illustrate that approaches to governing data to secure privacy protection differ considerably because they are tailored to the concerns and inequalities in specific political and economic contexts. It is therefore important to differentiate between approaches in the

⁴⁸ Bryant (2021).

⁴⁹ Andere & Kathure (2024); Ndemo & Thegeya (2023); South Africa Government (2024).

⁵⁰ Government of Brazil (2018).

⁵¹ He (2023); US-China Commission (2022); Voss & Pernot-Leplay (2024).

⁵² Chin (2024)

⁵³ Government of India (2023, p. 1); *Justice K.S. Puttaswamy (Retd.) & Anr. vs. Union of India & Ors* (2017); for an overview of trends in data governance, see also Punia et al. (2022).

⁵⁴ Abdulrauf & Dube (2024); Coos (2022) provides a comprehensive overview of data privacy laws in Africa.

higher-income countries and the middle- and lower-income countries.⁵⁵ Data protection and privacy legislation can share common themes, such as provisions for consent and attention to data subject rights, although this legislation differs in scope and specific provisions, such as data localization.

For instance, in the *European Union*, the GDPR mandates strict consent requirements and robust data subject rights, influencing global data protection standards with its comprehensive and extraterritorial reach. In contrast, in the *United States*, the CCPA, while also emphasizing consumer rights and transparency, introduces a unique private right of action and provisions tailored to California's legislative context. *India's* DPDPA incorporates stringent data localization requirements, reflecting an emphasis on digital sovereignty, local data control and geopolitical considerations. Each of these legislative approaches plays a role in shaping the practices of data privacy and individual rights protection and the standards that are adhered to.⁵⁶

Defining responsible parties. Under the GDPR, responsible parties are clearly defined as either 'data controllers' or 'data processors'. A data controller determines the purposes and means of processing personal data, while a data processor is responsible for processing data on behalf of the controller. This distinction is crucial for accountability as it clarifies who is responsible for ensuring compliance with the GDPR, and who will be liable if something goes wrong. In India, the DPDPA introduces the concept of data fiduciary and data processor, where data fiduciaries are akin to data controllers under the GDPR and are tasked with exercising due diligence in the processing and securing of personal data.

Governing data is a delicate endeavor for policy makers. It involves a struggle to manage the balance between the need for data security and privacy and the benefits of data utilization (for sector applications, e.g., health, finance, environment monitoring or for monetization purposes). By implementing specialized governance measures, enhancing transparency and promoting public awareness, the aim is to safeguard sensitive personal data against misuse. The diverse parties involved in data collection – from tech giants and startups to governments and third-party contractors – presents unique challenges in achieving effective governance that can assure accountability, fairness and transparency in how data is collected and used.⁵⁷

4.3 GOVERNING DIGITAL PLATFORMS

Both anticipatory and remedial forms of governance are being applied in many countries to moderate the behavior of the big tech companies, with measures being put in place to establish rules for digital platform operation when it is found to be inconsistent with human rights standards and/or to be anti-competitive. This section highlights how these measures impact on the problem of mis- and disinformation, but does not address the full complement of governance measures being introduced in regions around the world.

The *European Union* introduced measures dedicated to countering mis- and disinformation with an Action Plan in 2018. This was centered around improving capabilities to detect, analyze and expose mis- and disinformation. The aim was to strengthen coordinated responses, mobilize the private sector, raise awareness and increase societal resilience. A Code of Practice on Online Disinformation was put in place (and strengthened in 2022). The Code commits industry to address mis- and disinformation, political advertising and the integrity of services, and aims to empower consumers and the research community.⁵⁸ It operates under the

⁵⁵ Sampath (2021).

⁵⁶ Park (2020).

⁵⁷ Dolata et al. (2022).

⁵⁸ EC (2018, 2022e); early signatories were Facebook, Google, Twitter and Mozilla, and parts of the advertising industry, followed by Microsoft and TikTok; see also Saurwein & Spencer-Smith (2020). For criticisms of the Code, see Culloty (2021); Monti (2020); Nenadić et al. (2023); Pamment (2020).

supervision of the European Commission, and has inspired other countries to take similar action.⁵⁹ A permanent monitoring mechanism – the European Digital Media Observatory – was established as a hub for fact-checkers and for those studying mis- and disinformation issues.⁶⁰

A wider European Union regulatory framework has been put in place to strengthen digital governance, including the Digital Services Act (DSA) and the Digital Markets Act (DMA).⁶¹ These rules share a foundation in human rights aiming to safeguard citizen rights, and this legislative package is shaping content governance approaches at the national level.⁶²

Horizontal human rights. The DSA sets limits to the terms and conditions of platforms to govern the interactions between users and platforms – including the degree to which algorithmic recommender systems are used. Article 14 mandates that platforms must consider user interests in content moderation and complaint handling, referencing fundamental rights, such as freedom of expression. Very large online platforms (VLOPs) are required to respect fundamental rights due to the ‘systemic risks’ they present, based on comprehensive risk analyses and mitigation strategies. Article 34(1) obligates platforms to consider users’ fundamental rights among other factors when evaluating risks.⁶³

The DSA is ‘a horizontal framework for regulatory oversight, accountability and transparency of platforms and search engines’.⁶⁴ Many of

its measures apply to digital platforms and intermediaries with more than 45 million users per month in the European Union.⁶⁵ The Act’s provisions govern the algorithms used in automated content moderation, with binding obligations to remove illegal content, safeguards to respect freedom of expression and substantial penalties for failure to comply. VLOPs and search engines must adhere to a benchmark for processing valid notifications for removal of illegal hate speech in less than 24 hours. If a platform considers that content is not compliant with its terms and conditions, it may proceed with deletion or restriction. The aim is to create a safer digital space within which the fundamental rights of all users of digital services are protected.⁶⁶ These legislative measures combine ‘internal market, fundamental rights and geopolitical motivations, primarily in relation to emerging technologies’.⁶⁷

The DMA addresses the monopolistic behavior of the largest digital platforms with the aim of establishing a ‘level playing field’, that is, a contestable market, by constraining the practices of companies with gatekeeping power and that offer ‘core’ platform services. The overall goal is to promote ‘innovation, high quality of digital products and services, fair and competitive prices, as well as high quality and choice for end users in the digital sector’⁶⁸ by addressing imbalances in bargaining power and unfair (monopolistic) practices so that greater choice is available to platform users. There are sanctions against platform self-preferencing, the largest gatekeepers must enable the interoperability of services, and there are other measures aimed at achieving a balance between business and individual (or collective) interests.

⁵⁹ DiGi (2022); Wilding (2021).

⁶⁰ European Digital Media Observatory: <https://edmo.eu>.

⁶¹ EC (2022a, c). For the Data Governance Act, see EC (2022d) and for the Data Act, see EC (2023); see also Akman (2022); Botta (2021); Broughton Micova & Jacques (2020); Galantino (2023); Just (2022); Mansell (2021); Moreno Bellosso & Petit (2023); Nenadić *et al.* (2023).

⁶² Church & Pehlivan (2023), authors affiliated with Linklaters, a law firm, with offices in London and Madrid. The large tech companies are also subject to national law with binding measures, such as the German Network Enforcement Act 2017, the French Organic Law No. 2018–1201, and Hungarian legislation; see German Law Archive (2017); Government of France (2018); Stolton & Maksimov (2020).

⁶³ Defined as platforms with more than 45 million users per month.

⁶⁴ Nenadić *et al.* (2023, p. 8).

⁶⁵ Turillazzi *et al.* (2023).

⁶⁶ EC (2022c); Reyna (2024).

⁶⁷ Broeders *et al.* (2023, p. 1272), funded by the Dutch Ministry of Foreign Affairs (BZ, Ministerie van Buitenlandse Zaken); see also Mansell (2021).

⁶⁸ EC (2022a, para. 106). See also Brown & Marsden (2023); Crémer *et al.* (2019).

Competition/anti-trust legislation provides tools that are applied *ex post* to mitigate harms. Remedial remedies can include corporate divestment, fines and behavioral requirements. Competition law is seen as a means of leveling the market and diffusing gatekeeper power, although the gatekeeping power of big tech companies is generally treated as a ‘natural’ outcome of technological innovation.

Competition law applies in the European Union, and cases have been brought against Google’s search and advertising practices, Google’s and Apple’s app store rules for participation, Meta’s data collection and processing practices, and Amazon, for its treatment of companies that use its online marketplace.⁶⁹ The Treaty on the Functioning of the European Union (TFEU) applies to the conduct of ‘gatekeepers’. The scope tends to be limited to cases where the dominance of specific markets can be evidenced through lengthy investigation, although the criteria for establishing market dominance are slowly being modified.⁷⁰ For example, in Germany, non-price issues, such as access to data, have been treated as a potential criterion for determining market power, and member states are introducing modifications to enable them to bring actions against digital platforms more easily.⁷¹

The digital platforms have faced few efforts in the *United States* to curtail their market power until recently, allowing them to refine their business models to maximize user engagement and monetize data for profit. This has enabled them to acquire or suppress competitors, favor their own products and services, and downplay or disavow responsibility for harms linked to data collection, processing and monetization operations. The companies insist that they are providing their customers with convenient ways to access digital content and to buy goods online consistent with their individual preferences. However, more aggressive application of anti-trust law was encouraged under the Biden Administration,

with cases being brought against the platforms by the Department of Justice and the FTC as they pursue more vigorous efforts to limit platform monopolistic behavior.⁷² Proposals for sector-specific legislation, with some echoes of European Union approaches, are considered from time to time at the federal level to tackle big tech power. These have not been signed into law, but they call for prohibitions on large platforms giving preference to their own products, encourage interoperability and restrict platform use of non-public data, with penalties and injunctions.

In the United States there is much debate about the spread of viral mis- and disinformation and the consequences of content governance practices.⁷³ The First Amendment speech rights protections have led to controversy around the need for content governance.⁷⁴ Legislative proposals aimed at curtailing the circulation of content deemed to be harmful typically fail to attract sufficient congressional support. Digital platforms benefit from Section 230 of the Telecommunications Act of 1996.⁷⁵ Providers or users of interactive computer services are not treated as ‘publishers’ or ‘speakers’ of any information provided by another content provider. They therefore have broad immunity from liability for the content they host. Debates about how platform immunity might be circumscribed are highly politicized. Proposals to combat ‘fake’ information are met with ‘free market’ arguments and the claims that competition will eliminate problems.

This report does not cover all the cases seeking to curtail the big tech companies’ power. However, it is important to note that when there are successful cases confirming their monopoly power, this could have a substantial long-term impact. One example is a court ruling in August 2024 that Google was a monopolist in the general search text ad market. However, it was not found to be a monopolist in the search ads market, that is, based on the signals

⁶⁹ Nicoli & Iosifidis (2023).

⁷⁰ EC (2012, Articles 101, 102).

⁷¹ Just (2018, 2022).

⁷² FTC (2024); see also Stigler Committee (2019); Wu (2018).

⁷³ Flew (2021).

⁷⁴ Forum on Information and Democracy (2024d).

⁷⁵ US Congress (1996b).

provided by users' online interactions and the company's algorithms.⁷⁶ At the time of writing the judgment was under appeal. If it stands, it could open the door to further action, from breaking up monopolies to forcing companies such as Meta, Apple and Amazon to change their behavior, for example to modify their algorithms or make them provide support to the news media industry.⁷⁷

None of these judgments changes the overarching commitment to rapid innovation in digital technologies, including the use of opaque algorithms and generative AI (GenAI) for profit, which, in the *United States* at least, remains a powerful mobilizer of investment in future generations of data monetization strategies.⁷⁸ An argument gaining some ground is that the scale of digital platform adoption has reached a point where they have become essential public services and should be subject to the same regulations as public utilities (as privately, publicly, cooperatively or municipally owned) that operate as 'natural monopolies'. However, whether digital platforms such as Meta meet the threshold for being classified as an 'essential service' is disputed, and some argue that treating them in this way could entrench their monopolistic position.

Much of the literature on big tech governance focuses on the *United States* and *Europe* (and increasingly on *China*). Other countries also have legislation. We mention only a small sample of instances here, where measures are being taken to combat mis- and disinformation.⁷⁹

Country measures to legislate to limit mis- and disinformation. Between 2011 and 2022, 78 countries had passed sector-specific

laws designed to limit the spread of online mis- and disinformation. Some focus on improving transparency and accountability and increasing media and information literacy. Others focus on criminalizing the creation and distribution of content, which, in authoritarian states, paves the way for subjective evaluations of what constitutes 'fake news', leading to the abuse and undermining of freedom of expression, including press freedom.⁸⁰ National legislation aimed at combating mis- and disinformation includes the *Malaysian Anti Fake-News Act 2018*, the *Singapore Protection from Online Falsehoods and Manipulation Act 2019*, the *Russian fake news law*, the *Bangladesh Digital Security Bill Act*, and several laws in *China*.⁸¹

These laws tend to position digital platform owners or states as arbiters of 'truth', which can lead to abuses of basic freedoms.⁸² Legal initiatives also face lobbying by the big tech firms. For example, in *Brazil*, work on a law on AI initially proposed in 2019 had not been adopted at the time writing in late 2024 due to successful lobbying by big tech companies. The draft calls explicitly for the 'development, implementation and use of Artificial intelligence in *Brazil* ... based on integrity of information through the protection and promotion of reliability, accuracy and consistency of information'.⁸³ In authoritarian regimes, when digital platforms provide spaces for political activism – including by opposition parties – this is problematic from a rights-based perspective, and is illustrated by the experience of *Southeast Asian* states, where state authority is maintained through a combination of political pressure and internet controls.⁸⁴

⁷⁶ This general search ad market excludes display ads, retargeted display ads and non-search social media ads, that is, ads that rely on 'indirect signals to decipher a users' latent intent' based on a user's past online interactions (US District Court, 2024, p. 168).

⁷⁷ Radsch (2024).

⁷⁸ This is addressed in Chapter 8.

⁷⁹ Pickard (2022a).

⁸⁰ Lim & Bradshaw (2023).

⁸¹ Dittrich (2019); Malaysia Government (2018); Repnikova (2018); Reuters (2019); Richter (2019); Singapore Statutes (2019).

⁸² Dittrich (2019).

⁸³ Government of Brazil (2023); our translation.

⁸⁴ Sinpeng (2020).

4.4 GOVERNING AI SYSTEMS

Regulators and policy makers face substantial challenges in defining rules for content governance enabled by AI systems in the light of the challenge of balancing the potential benefits to be gained from encouraging innovation against the risk of harm to individuals, businesses and society from the lack of regulation to protect them.⁸⁵ International bodies such as the Council of Europe, the OECD and the United Nations, and its agency, UNESCO, are active in defining principles and standards designed to protect human rights against the negative impacts of AI systems.⁸⁶ Every intervention intended to uphold human rights norms is therefore a balancing act that must be assessed in each context.

Differences in approaches are apparent in AI governance initiatives announced by the United States and the European Union in 2023. President Biden's 2023 Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence emphasizes the obligations of AI developers of 'dual-use' foundation models to show that these will not lead to violations of federal laws on civil rights, discrimination, etc.⁸⁷ In contrast, the European Union's AI Act of 2024 takes a wider view that includes obligations on the part of AI developers to actively protect human rights.⁸⁸

Before the emergence of AI-related regulation, the components of information ecosystems were already regulated at various levels: international law, regional standards and national laws. Recent initiatives to regulate the impact on AI on societies have started to home in on transparency requirements, training data disclosures and risk assessment obligations. Normative approaches

include the United Nations resolution, 'Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development', the OECD's Recommendation of the Council on Artificial Intelligence, UNESCO's Recommendation on the Ethics of AI, the G20 AI Principles, the G7 Hiroshima Process, including principles for GenAI, the AI Safety Summit Declaration in Bletchley, the Council of Europe Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, and the Executive Order in the United States. The European Union's AI Act enters into force in 2025.⁸⁹ Many of these aim to guard against the risks and harms of mis- and disinformation.

Policy makers and their regulatory institutions face two connected challenges when developing rules for governing the impacts of AI on information ecosystems.⁹⁰ First, 'AI' is not a static product that can be regulated once, and regulatory approaches need to focus on the evolution of AI systems during their whole lifecycle, that is, throughout the design, development and deployment phases. Second, the impacts of the use of AI systems are seen as being more determinative of regulatory needs than abstract characteristics of a system (which are bound to change). For this reason, most regulatory approaches involve risk-based approaches that are used to define AI systems requirements based on the level of risk a system is judged to pose.⁹¹ The aim of risk-based approaches is therefore seen by some as 'not primarily to manage risk but instead to ensure legislative proportionality' that will avoid stifling innovation.⁹² An example of this is the European Union's AI Act of 2024, which classifies AI applications into different risk categories, with more extensive obligations for higher-risk applications.

⁸⁵ For reviews of research on the governance of algorithms and AI, see Gritsenko *et al.* (2022), supported by NOS-HS (Joint Committee for Nordic research councils in the Humanities and Social Sciences); Crawford (2021); and for resources on legal approaches, see Custers & Fosch-Villaronga (2022); De Bruyne & Vanleenhove (2021); Książak & Wojtczak (2023); see also Bullock *et al.* (2022).

⁸⁶ Bello y Villarino (2023). Professional societies, such as the Institute of Electrical and Electronics Engineers (IEEE) are very active in this domain, especially in relation to Electronic Warfare (EW). See, for example, Koene *et al.* (2018).

⁸⁷ US Executive Order (2023, para. k). The dual-use foundation model is defined as an 'AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters.'

⁸⁸ European Commission (2024b); Larsen & Küspert (2024).

⁸⁹ Council of Europe (2024); G7 (2023); G20 (2019); OECD (2022c); UK DSIT (2023); UNESCO (2022c); UN (2024c); see also EC (2024c), agreed March 2024; US Executive Order (2023).

⁹⁰ De Gregorio (2023).

⁹¹ Cole (2024).

⁹² Mahler (2022, p. 247).

AI system risk categories. In the *first category*, certain AI systems are deemed so risky as to be ‘unacceptable’: an AI-based Assessment of individuals’ behavior by government agencies, that is, when they influence the ‘free will’ of users or contain ‘social scoring’. Under the AI Act’s scope, their use is prohibited in the European Union.

The *next category* includes ‘high-risk’ AI systems, which are listed in Annexes II and III of the Act. Annex II features a list of existing European Union regulations that require a ‘conformity assessment’ for products that bear specific risks. If an AI component is part of these products or the product ‘itself’, it is considered a ‘high-risk’ AI system. For the list in Annex III, the context of use is more relevant, that is, it is not the AI system itself that is considered risky, but the domain in which it is applied. Eight domains are named in which certain AI systems are ‘high risk’, such as those involved in decisions about access to education or employment. A particularly large number of applications that are considered ‘high risk’ are those involved in law enforcement or migration. If an AI system falls into this category, manufacturers and users must adhere to compliance obligations, such as having risk governance and quality management systems in place, and registering the AI system with the European Commission.⁹³

The *third category* includes ‘low-risk’ AI systems, for which the Act requires ‘only’ transparency obligations and thus, significantly fewer requirements than for those in the ‘high-risk’ category. This means that providers of AI systems that (1) interact with humans, (2) are used for emotion or biometrics recognition, or (3) that generate ‘deepfakes’ must notify their users that the content was generated by AI.

Not regulated by the AI Act are ‘risk-free’ AI systems that include, for example, spam filters for email programs. Here, the risk for users is considered so small that no regulation is envisaged. For emerging AI systems not previously addressed, the AI Act stipulates that they must be categorized as ‘high-risk’ AI systems if they can negatively affect fundamental rights. This classification imposes substantial compliance duties on both providers and users of these AI systems. Article 13 requires providers of ‘high-risk’ AI systems to transparently outline the risks these systems pose to fundamental rights when employed, and Article 14(2) mandates human oversight of ‘high-risk’ AI systems to safeguard fundamental rights.

The AI Act also references fundamental rights at various points. These often serve to clarify the broader context and rationale for specific provisions at European Union level, highlighting the potential of AI systems to impact fundamental rights. For instance, recitals (legislative texts) address the risks of AI systems being used for manipulative or exploitative practices. One criticism voiced against regulating only primary uses of AI models, mainly exercised through quality assurance of their training data, is that risks of secondary use, where a model used in an AI system is applied in a way that its developers did not intend, may go undetected. A solution would be to focus on the concept of purpose limitation for AI models, which would leverage existing data protection approaches.⁹⁴ Some researchers are encouraging more stringent AI systems rules, arguing that law makers should learn from both ineffective and missing regulations during the early days of social media, when they failed to address the underlying business model that led platforms to prioritize the data-driven monetization of user attention.⁹⁵

Many other regions and countries are putting strategies and governance frameworks in place.⁹⁶ Despite the fact that AI systems are largely developed by companies in the Global North and China, organizations in African countries, Latin America and Asia are

⁹³ Annexes I and III of the Act refer to harmonization with European Union legislation and listed high-risk AI systems requiring third-party conformity assessment.

⁹⁴ Mühlhoff & Ruschmeier (2024).

⁹⁵ Sanders & Schneider (2024); and there are calls for standards, see Lewkowicz & Sarf (2024); Schwartz *et al.* (2022), although some argue that companies will use voluntary standards to evade regulations on AI systems development (Han *et al.*, 2022).

⁹⁶ For a comprehensive review of AI systems regulation in emerging economies, see Findlay *et al.* (2023).

developing applications using large data sets and machine translation tools, and there are calls for the localized development of AI applications.⁹⁷

- The *African Union* agreed a Continental Artificial Intelligence Strategy in July 2024. A review of the state of AI regulation in Africa in 2024 indicates that AI governance measures face challenges of ‘weak institutional frameworks, limited judicial capacity, lack of expertise from policymakers, fragmented laws, and poor enforcement mechanisms, where laws, even if existing, are seldom applied’.⁹⁸
- There are calls to reframe debates about AI governance in Global Majority regions to acknowledge power asymmetries and to recognize that the aim to develop ‘responsible AI’ governance frameworks still allows powerful companies to ‘diffuse accountability, evade liability, and disregard rights’.⁹⁹
- Discussions around AI governance typically exclude ‘marginalized communities and groups including women, racial and sexual minorities, small producers, workers, and Indigenous communities’.¹⁰⁰
- It has been pointed out that debates around ethical issues and requirements to ensure that AI systems are ‘explainable’ are rooted in Western perspectives – for example, in *sub-Saharan African countries*, where local informal savings and lending practices are common, AI tools to assess creditworthiness exclude these practices.¹⁰²
- In *China*, some argue that academic input into shaping AI regulation is considerable, and that the emphasis is on strong binding regulations.¹⁰³

Clearly uniform approaches to governing AI systems and tackling mis- and disinformation are not viable.¹⁰⁴ And when AI governance turns to ethical considerations, scholars in Global Majority World countries point to the bias of debates towards the interests of the Global North, which neglect approaches that differ from those adopted in *Europe* or the *United States*.¹⁰⁵ Proposed AI legislation in *Brazil*, for example, follows the European Union’s AI Act in adopting a risk-based approach with a list of prohibited applications. It differs, however, in guaranteeing individual rights accompanied by judicial and administrative mechanisms to enforce these rights. These include the right to contestation and human intervention, emphasizing due process for people affected by automated decisions.¹⁰⁶

Frameworks are being developed that transcend national boundaries and address the international implications of AI systems, beyond regional normative approaches and global commitments to sustainable and accountable AI. However, so far no organization has succeeded in taking the lead in driving the development of AI systems in a way that is based on international solidarity and inclusive participation.

4.5 GOVERNING NEWS MEDIA

Rule-based governance arrangements have implications for the way the news media is regulated, especially since what counts as news, what is a news media organization, and journalism profession norms and practices are changing, or at being least contested, in many countries.¹⁰⁷

Legacy and online news media are intertwined in the data economy. The governance of data, digital platforms and AI influences the health of

⁹⁷ Okolo (2023).

⁹⁸ African Union (2024).

⁹⁹ Tech Hive Advisory Center for Law & Innovation (2024, p. 21).

¹⁰⁰ Gurumurthy & Bharthur (2023, p. 2).

¹⁰¹ Gurumurthy & Bharthur (2023, p. 3).

¹⁰² Effoduh (2024).

¹⁰³ Zhu (2022), supported by the Finnish National Agency for Education.

¹⁰⁴ Kakkar (2023).

¹⁰⁵ Gunkel *et al.* (2024); the need to differentiate between countries is illustrated by a comparison of AI systems in Senegal and Cambodia (Heng *et al.*, 2022).

¹⁰⁶ Government of Brazil (2023); Mendes & Kira (2023).

¹⁰⁷ See Section 4.1, Chapter 2 for a discussion of changing journalism practices.

information ecosystems. Verifiable news and informed public opinion are essential for democracy to function, and this requires high-quality public debate and deliberation and accountable representation. When mis- and disinformation circulate and the digital platforms operate in ways that depart from human rights expectations, there is no doubt that this contributes to democratic fragility.¹⁰⁸

The news media are expected to preserve and maximize diversity and a plurality of voices in the public sphere within the framework of internationally agreed rights and responsibilities. Yet news media outlets face the challenge of declining levels of trust, some people are actively turning away from the news, there is a deficit of media pluralism, a growing dependence of news media organizations on digital platforms, increasing concentration in the news media industry in many countries, and absent or weak editorial independence.

News media regulation can backfire when it is used as a pretext to consolidate state power and control over information flows, which leads to censorship and repression or more subtle forms of leverage that hold news media organizations in check.

- In *Cuba*, the state maintains control over the mass media (also dominating artistic and intellectual affairs) by prohibiting private (legacy and online) media outlets under the 2019 Constitution, which classes them as being funded by 'enemies of the state'.¹⁰⁹
- In *Hungary*, the use of media laws, efforts to control regulatory bodies and a concentrated media market have helped to consolidate domination by the ruling party.¹¹⁰
- When Apartheid ended in 1994 in *South Africa*, new governance arrangements for the media were introduced. The 1996 Constitution gave

unprecedented levels of freedom to media organizations, emphasizing the priority to build an ethical, independent and publicly accountable news media, and moving from media self-regulation to co-regulation.¹¹¹ Yet there are complaints that the news media serves the interests of an elite, that disadvantaged community voices are not represented, and that the public broadcaster, South African Broadcasting Corporation (SABC), lacks independence. SABC is criticized for being too soft on the elected government, and debate focuses on whether the news media should be more critical of the democratic government or protect democracy by supporting it.¹¹²

- In *Turkey*, the government has sought to foster a favorable news media by leveraging structural, legislative and illegal measures to benefit the ruling party. After a failed coup in 2016, a restructuring of the media system led to greater repression through certain measures, including economic incentives, structural support for favoring the ruling party and control of regulatory bodies.¹¹³
- In *Venezuela*, a legal reorganization under the Hugo Chávez government (1999–2013) shifted the media system from private dominance (opposing the government) to state dominance (supporting the government), without alleviating the political and economic pressures on news media organizations. Under Nicolás Maduro (2013–19), the news media experienced further government pressures.¹¹⁴
- *Vietnam* and *Singapore* have implemented media regulations, including censorship, ownership controls, personnel management and other repressive instruments. Vietnam's approach is coercive while in Singapore, political norms are enforced implicitly by embedding stakeholders with financial interests in the media system.¹¹⁵

¹⁰⁸ Pickard (2022a, b); Tambini (2021); Tenove (2020).

¹⁰⁹ García Santamaría & Salojärvi (2020); Romeu (2023).

¹¹⁰ Polyák (2019).

¹¹¹ Wasserman (2020b).

¹¹² Wasserman (2020b).

¹¹³ Akser & Baybars (2023).

¹¹⁴ García Santamaría & Salojärvi (2020).

¹¹⁵ Haenig & Ji (2024), supported by the National Social Science Fund of China.

Repressive measures infringe on human rights, and they also create a space to produce mis- and disinformation and its circulation through both legacy and online news media.

In addition to political pressure, many news media organizations are dependent on digital platforms to circulate their news, and many news organizations are facing financial pressures, which leads to questions about their independence.¹¹⁶ Declining advertising revenues can prevent news media organizations from fulfilling their democratic function, reduce media pluralism and contribute to perceptions that the news media industry is untrustworthy.

Power asymmetries between news media organizations and the big tech owners of digital platforms are visible in multiple regions. There have been clashes among the platforms, news organizations and regulators in *Australia*, *Canada* and the *European Union*, for example.¹¹⁷ One remedy is to compensate publishers for the content that platforms host, since platforms derive substantial economic value from featuring news on their sites, although this can lead to the largest news media organizations benefiting disproportionately.¹¹⁸ There are also disputes about the scale of compensation, especially among economists who argue that the digital platforms do not ‘free ride’ on the news media, and that payments by platforms to the publishers would inhibit innovation, among other reasons.¹¹⁹ Other means of financially supporting news media, such as introducing taxes on digital advertising, are also being proposed.¹²⁰

In the Global North and Global Majority World countries, there are moves to empower smaller news organizations to bargain collectively with big tech companies. Other measures include influencing the production, distribution and monetization of news content – for example, sometimes using their own apps on a subscription basis (mainly viable for the largest providers) by setting up paywalls or membership programs, creating their own real-time advertising marketing capability, or launching cooperative news organizations.¹²¹

Public service media (PSM) (including those permitted to attract advertising) are rarely economically sustainable without subsidies, concessions and/or protections that involve direct government financial support, license revenue, technical assistance and collaborative strategic programming and advocacy.¹²² If news media are treated as ‘a public good’, this can help to maintain independent PSM organizations.¹²³ In countries where PSM is reasonably shielded from political pressure, these news organizations are a vital component of a healthy information ecosystem.¹²⁴ In countries where governments pressure news media organizations, both privately owned outlets and PSM often fail to meet normative expectations.¹²⁵

In some countries and regions action is being taken to try to promote news media pluralism and media freedom and to counter mis- and disinformation.¹²⁶ For example, the *European Union* introduced a Democracy Action Plan in 2020, which included measures to promote free and fair elections, strengthen media freedom and counter mis- and disinformation.¹²⁷

¹¹⁶ Wasserman (2018); see also Chapter 2 for discussion of news media independence.

¹¹⁷ Marshall (2023); Meese & Hurcombe (2021), funded by the Australian Research Council (ARC); Hermida (2023); see Section 2, Chapter 2 for more details on news media concentration.

¹¹⁸ Flew (2023); Flew & Martin (2022).

¹¹⁹ Lesh (2023); in early 2024 it was estimated that the platforms in the United States would owe news publishers annually between USD 11.9 and 13.9 billion – the methodology is explained in Mateen *et al.* (2023), two authors affiliated with the Brattle Group, US.

¹²⁰ Radsch (2022).

¹²¹ Grover & Baik (2024); MacKenzie *et al.* (2023); Marshall (2023); Poell *et al.* (2023).

¹²² Radsch (2022).

¹²³ UNESCO (2022d).

¹²⁴ Michalis & D’Arma (2024).

¹²⁵ Farahat (2021).

¹²⁶ Paal (2017).

¹²⁷ EC (2020b).

The Media Freedom Act. In the European Union, the Media Freedom Act aims to protect journalists' work, secure the independence of public media and increase the transparency of private media ownership. It requires a fair allocation of state advertising revenue to news media producers and aims to secure media freedom. The Act obliges member states to implement media concentration assessments (although it neither prevents media concentration nor sets a threshold). Article 22 introduces a 'media pluralism test', requiring member states to examine media mergers based on the implications for media pluralism and editorial independence, as well as market competition assessments. This is a substantial shift away from the previous hands-off approach to regulating media pluralism.¹²⁸

corporate self-governance to co-regulation (state and corporate) to direct state intervention. All the components of information ecosystems, from the network infrastructure to the service applications layer, are subject to norms and rules that condition how they operate. These are expected to be consistent with broad principles, including for how data is collected and processed. We have presented the features of selected governance arrangements that are being put in place (and in some cases, resisted) around the world that influence information integrity and the health of information ecosystems.

There is broad agreement that states have a duty to act to protect human rights and fundamental freedoms. This includes a negative obligation not to violate rights – including those of big tech companies. States also have a positive obligation to protect human rights and implement them. This means that every aspect of governance involves a balancing act, with an outcome that varies with each context.

The synthesis of research in this chapter shows that:

- On the infrastructure layer of information ecosystems, network neutrality policies and 'zero-rating' regulations are central to how those who have connections and affordable access experience these ecosystems. These policies and others, such as internet shutdowns and social media blocking during elections or political unrest, contribute to fragmenting the internet and curbing access to information in many regions of the world. These policies and practices are informed by state ambitions to achieve digital sovereignty and corporate interests in profit.
- Governing how data is produced and used is increasingly controversial because of the lack of transparency in corporate data collection and monetization, targeted advertising and the

The news media in Western democracies have been largely self-governing to protect their independence. The freedoms enjoyed by the news media historically have never been absolute in any country, and the privileges and duties of the journalism profession have varied across the world. In response to changes in the relations between news organizations, the platforms and state actors engaged in producing and circulating news, it is essential that human rights standards provide guidance on normative expectations, even if there are deviations in practice.¹²⁹

5 Chapter Summary

This chapter has described the approaches applied by national governments (or regions) to govern the growing complexity of information ecosystems. These are spread across a spectrum, of hard and soft touch regulation: from voluntary

¹²⁸ EC (2024b, p. 3). The Act excludes user-generated content unless it is uploaded for financial or other consideration, purely private correspondence and services that do not have 'provision of programmes or press publications as their principal purpose', corporate communication and informational or promotional materials, but it includes freelancers. See also Brogi *et al.* (2023); Centre on Media Pluralism *et al.* (2022).

¹²⁹ Tambini (2021).

misuse of data for political gain. Legislation, such as the European Union's GDPR, is not a panacea for all data economy issues. It is important to attend to how approaches in higher-income and middle- and lower-income countries differ.

- Approaches to governing big tech-owned digital platforms, such as the European Union's Digital Services and Digital Markets Acts (and codes of practice to counter mis- and disinformation), have achieved prominence in debates about how to limit the spread of online mis- and disinformation. These place obligations on the largest platforms to take down illegal or suppress harmful mis- and disinformation. Governance measures vary significantly around the world regarding the penalties or criminalization of those who produce and circulate mis- and disinformation.
- AI systems governance focuses on balancing the potential benefits to be gained from encouraging innovation against the risk of harm to individuals, businesses and society from a lack of regulation. The European Union's AI Act shows how AI applications can be classified into risk categories, but homogeneous approaches to governing AI systems and tackling mis- and disinformation are unlikely to be viable. Frameworks are being developed that transcend national boundaries aimed at increasing transparency and accountability. So far, no organization has succeeded in taking the lead in driving the development of AI systems that are based on international solidarity and inclusive participation.
- Verifiable news and informed public opinion are essential if the public sphere is to provide a space for democratic participation. Regulatory measures applied to legacy and online news media can backfire when they are a pretext to consolidate state power and control information flows, leading to censorship or leverage over news media organizations. While news media freedom has never been absolute, and the privileges enjoyed by journalists and news media organizations vary throughout the world, human rights principles should guide normative

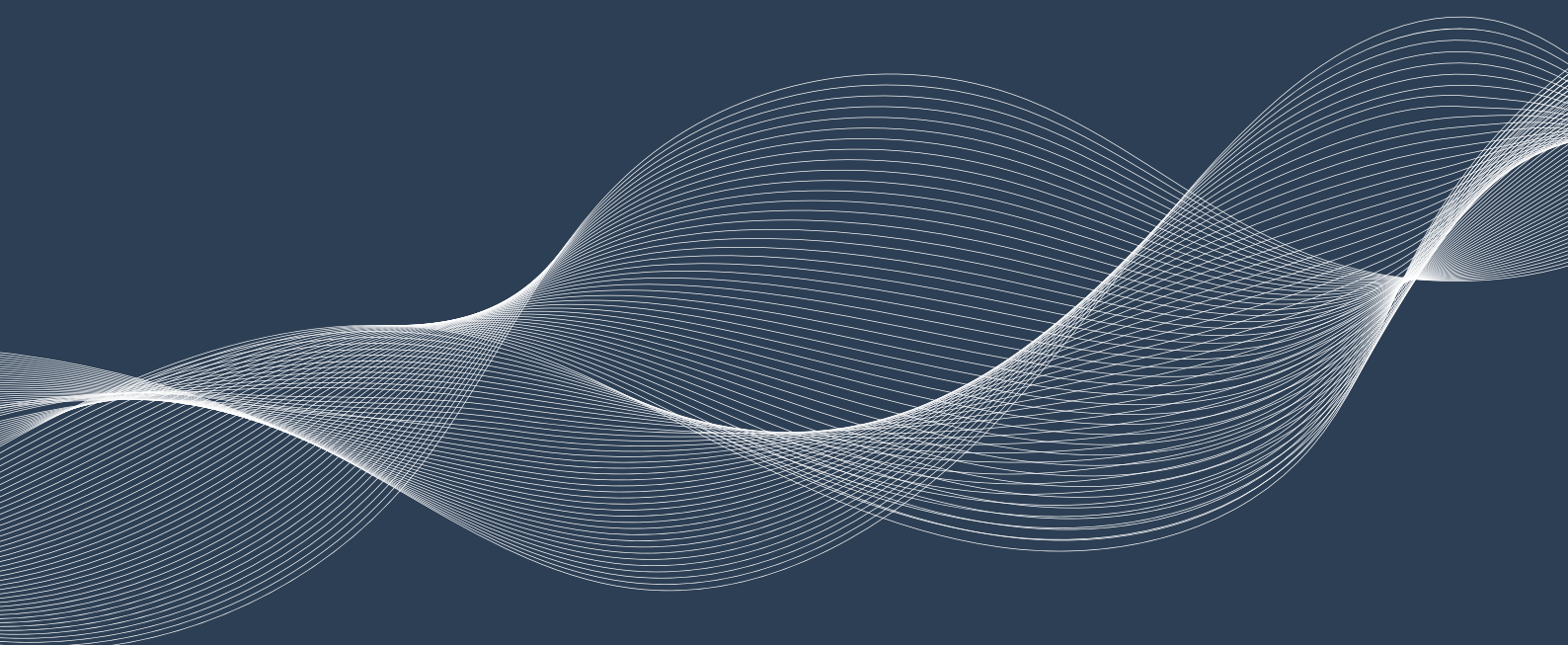
expectations, even when there are deviations in practice.

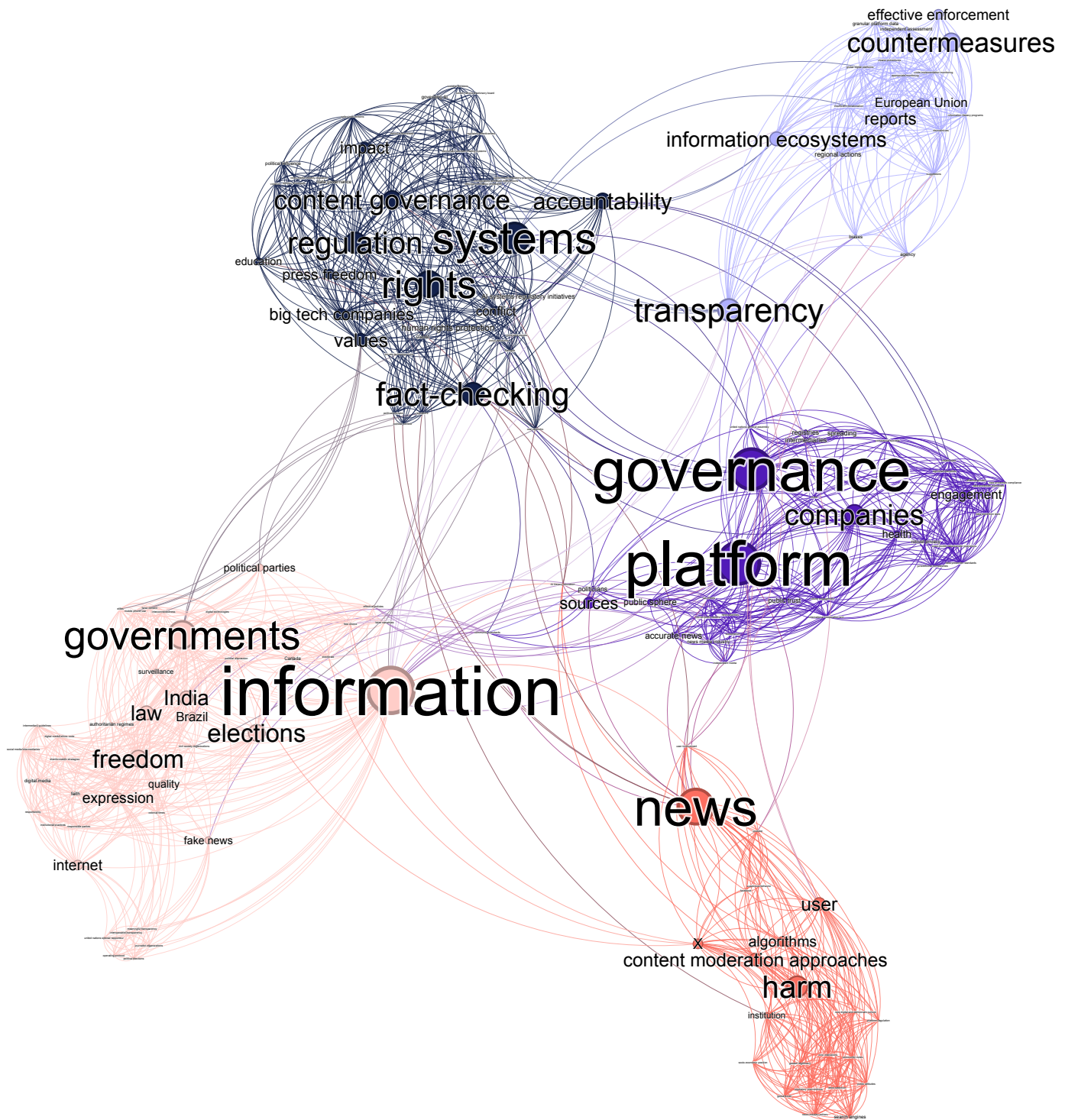
Research is needed:

- To monitor the voluntary and anticipatory or remedial governance measures that are being introduced globally in response to the strategies and practices of big tech companies, and to systematically track corporate lobbying that frames governance in these companies' interests.
- To monitor the implementation of governance measures, whether they uphold fundamental human rights and whether they are effective in helping people navigate information ecosystems to be resilient to mis- and disinformation. It is essential to differentiate between normative goals and principles being articulated on a global level, and how these are translated into practice at local, country and regional levels over time.
- To assess the implementation of network neutrality policies in different contexts and their consequences.
- To examine how specific types of customer contracts restrict people's ability to access information and to participate in an informed way in information ecosystems.
- To examine systematically and on an ongoing basis the extent to which privacy and data protection, platform regulation, AI systems and news media governance are aligned with individuals' interests and the collective interest. Research must be inclusive of the experience of the Global Majority World.

CHAPTER 7

COMBATING MIS- AND DISINFORMATION IN PRACTICE





This map represents a statistical summary of the thematic content of this chapter. The network graph represents relations between the words in the chapter, placing them closer to each other the more they are related. The bigger the node, the more present the word is, signalling its role in defining what the report is about. The colors represent words that are closely related to each other and can be interpreted as a topic.

The map is generated by the OID on the basis of the chapter's text using GarganText – developed by the CNRS Institute of Complex Systems. Starting from a co-occurrence matrix generated from chapter's text, GarganText forms a network where words are connected if they are likely to occur together. Clustering is conducted based on the Louvain community detection method, and the visualization is generated using the Force Atlas 2 algorithm.

[Link to the interactive map here](#)

This chapter looks in detail at specific governance measures to combat mis- and disinformation by civil society organizations and governments.¹

The research synthesis in this chapter focuses on:

- **What content governance efforts are being made to combat mis- and disinformation?** This focuses on the work of not-for-profit fact-checking organizations, and on big tech self-regulatory and state or co-regulatory measures aimed at strengthening the integrity of information. It examines approaches to countering mis- and disinformation in politics, AI systems development, the news media industry and privacy and data protection.
- **What are the challenges in achieving effective governance of information ecosystems?** This examines how AI systems can be governed to achieve accountability and transparency, how press freedom and the independence of news media can be secured, and the difficulties of governing corporatized information ecosystems.
- **In what ways are human rights protections jeopardized by governance aimed at curtailing online mis- and disinformation?** Problems in differentiating between legitimate and illegitimate speech online, and the extent to which human rights principles that countries sign up to are adhered to, are examined.
- **What is known about the public's appetite for interventions to moderate online mis- and disinformation?** This discussion explains why there is no 'silver bullet' to cultivate trust in information generally, and in news media specifically.

The discussion in this chapter emphasizes the need to differentiate between the stated aims of governance initiatives and their consequences when practice falls short of normative expectations.

Chapter 8 critically examines alternative data governance practices aimed at resisting injustices, biases and harms arising in big tech-dependent and data-intensive societies.

¹ For background reading, see Catalán-Matamoros (2023); Enghel & Noske-Turner (2018); MacCarthy (2023); McGuigan (2023); Montero & Finger (2021); Moore & Tambini (2018, 2021); Tan (2020); Tyagi *et al.* (2024); West & McAllister (2023). See Appendix: Methodology for details of literature review process.

1 Introduction

The governance of information ecosystems is framed by legal and institutional arrangements established globally, regionally and nationally by states, and by their interactions with corporate rules and practices through industry self-regulation, co-regulation and direct state intervention. The norms and practices of state actors and not-for-profit organizations vary, even when they are framed by commitments to international human rights agreements. As information ecosystems change, and mis- and disinformation are seen as a major problem, measures are being discussed and implemented with the aim of improving the health of information ecosystems and the quality of debate in the public sphere.² The question of whether these measures are consistent with human rights commitments or facilitate the uses of digital platforms in ways that enable the big tech companies and some states to infringe on these rights is the issue at the center of this chapter.

Any assessment of the effectiveness of efforts to combat mis- and disinformation should include an examination of the roles of not-for-profit organizations in fact-checking, platform self-regulatory practices and co-regulatory and state interventions in content governance. These are examined in Section 2. In Section 3 some of main challenges to effective information systems governance are discussed, focusing on governing AI systems, efforts to achieve press freedom and

the independence of news media, and privacy and data protection regulations. Section 4 examines the extent to which human rights protections are jeopardized by content governance measures aimed at curtailing online mis- and disinformation. Finally, Section 5 reviews evidence on the public appetite for interventions by companies and states to moderate online information.

2 Assessing Measures to Combat Mis- and Disinformation

Efforts to inform the public about the authenticity and credibility of news are becoming more common, along with efforts to mount counternarrative campaigns to mis- and disinformation. There is a large mix of efforts and an outpouring of reports on the strategies and their effectiveness.³ Some of these measures involve the platforms, but many involve civil society actors and, in some instances, government action. Table 7.1 gives an overview of a set of countermeasures, how much is known about them, whether they are effective and whether they scale easily. Although the study on which this table is based has an international focus, most of the cited evidence is by United States-based scholars, with some exceptions. The authors conclude that ‘there is no silver bullet or “best” policy option’.⁴

² See Chapters 6 and 7 for discussion of a selection of governance arrangements that impact on mis- and disinformation.

³ See Mendoza *et al.* (2023) and reports including those from the Center for AI and Digital Policy (CAIDP, 2022) covering 75 countries, updating information about the bearing of intergovernmental, country and regional policies on human rights protections (mainly privacy and freedom of expression) and democratic values.

⁴ Bateman & Jackson (2024, p. 2), Carnegie Endowment for International Peace, independent think tank supported by the Special Competitive Studies Project (SCSP), private foundation focusing on industry competitiveness, US.

Table 7.1

Selected mis- and disinformation countermeasures

Type of Action	Intervention	How much is known?	How effective does it seem?	How easily does it scale?
Public information	Supporting local journalism	Modest	Significant	Difficult
	Media literacy education	Significant	Significant	Difficult
	Fact-checking	Significant	Modest	Modest
	Labeling social media content	Modest	Modest	Easy
	Counter-messaging strategies	Modest	Modest	Difficult
Government action	Cybersecurity for elections and campaigns	Modest	Modest	Modest
	Statecraft deterrence and disruption	Modest	Limited	Modest
Platform action	Removing inauthentic asset networks	Limited	Modest	Modest
	Reducing data collection and targeted ads	Modest	Limited	Difficult
	Changing recommendation algorithms	Limited	Significant	Modest

Source: Bateman & Jackson (2024, p. 5).

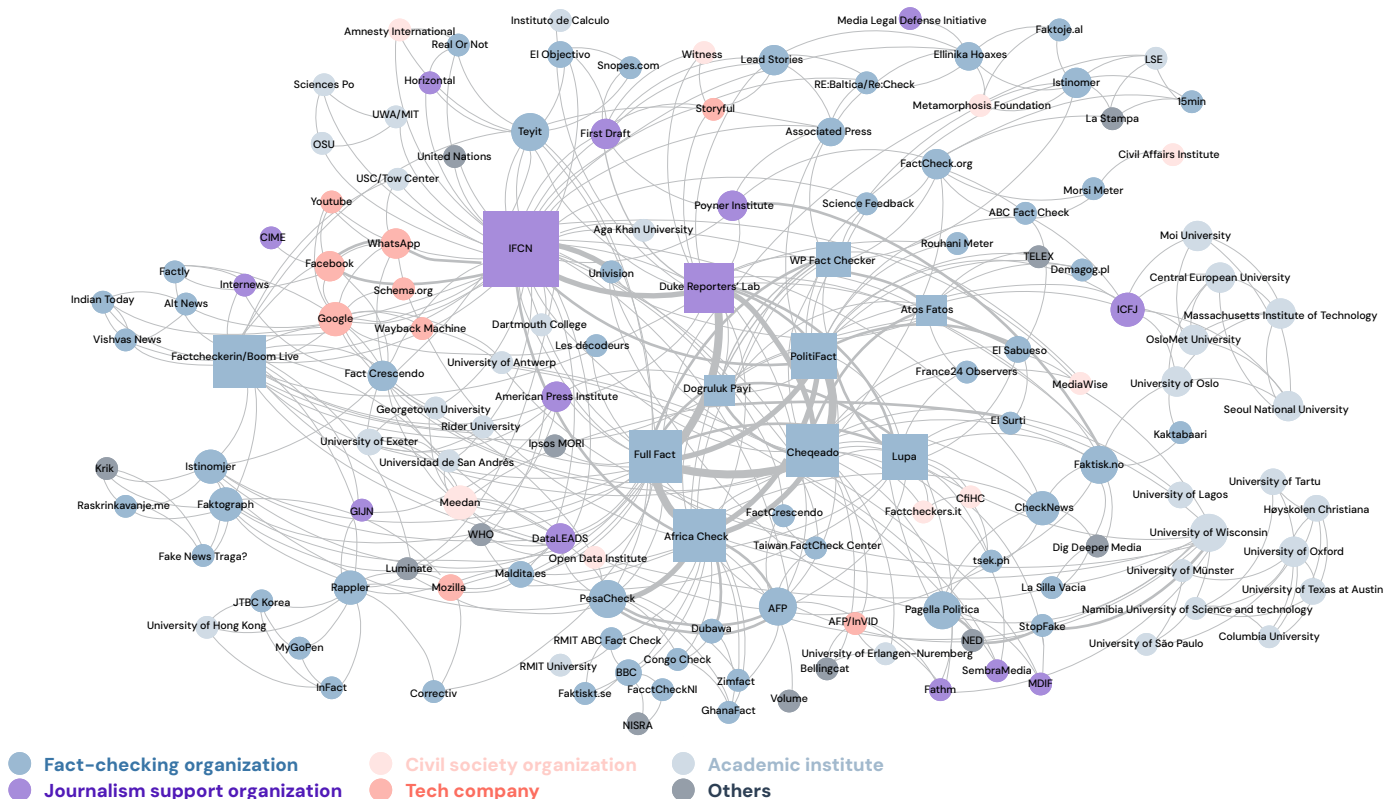
In earlier chapters, we discussed measures to support (local) journalism (Chapter 2) and media and information literacy (MIL) and AI literacy education (Chapter 5, where we explained that effects depend on the strategy), and commented on social media content governance, counter-messaging, content removal strategies and measures to change how algorithms operate. Here, the focus is on fact-checking strategies and measures to reduce data collection or to change how it is collected.

2.1 FACT-CHECKING ROLE IN TACKLING MIS- AND DISINFORMATION

Fact-checking was launched by a number of non-governmental organizations (NGOs). Since then, the number of third-party organizations offering fact-checking services has grown rapidly in response to the increasing volume of mis- and disinformation. By 2024 there were estimated to be 430 fact-checking organizations across more than 100 countries. They have developed partnerships with major tech platforms and built a transnational network, which remains fragile.⁵ Figure 7.1 shows the network of fact-checkers, as of 2021.

⁵ Graves & Cherubini (2016). According to Duke Reporters' Lab, there were more than 430 fact-checking organizations operating in 53 countries in April 2024. See <https://reporterslab.org/fact-checking>; Lauer & Graves (2024) report that Meta was paying fact-checkers in 119 countries in 2024 to identify misinformation on Facebook and Instagram. In a study supported by the European Research Council (ERC) and Horizon grant, Guo *et al.* (2022) survey how automatic fact-checking operates.

Figure 7.1
Network of fact-checking organizations, 2014–21



Source: Lauer & Graves (2024, p. 9).⁶

Fact-checking, defined as ‘the practice of systematically publishing assessments of the validity of claims made by public officials and institutions with an explicit attempt to identify whether a claim is factual’, is becoming common.⁷ In principle, fact-checking is like expert panel content moderation, but it maintains user agency by leaving the user to determine its relevance. As fact-checking organizations struggle to cope with the growing volume of ‘checkworthy’ claims appearing on social media platforms, there is increasing reliance for many on machine learning (ML) techniques.⁸ While many fact-checking organizations are committed to a process that retains a ‘human-in-the-loop’,⁹ automated fact-checking may be inevitable in the future, given the continuing rise in the volume of mis- and disinformation.

Evidence of the impact of fact-checking suggests that it has a limited effect on people’s responses to mis- and disinformation. Two surveys in the *United Kingdom* of people’s understanding of, and engagement with, fact-checking found that this process has a minimal impact on people’s news consumption. The study also found that participants express a desire for more television news fact-checking, but that the credibility of fact-checking in this context tends to reflect people’s beliefs in the impartiality (or otherwise) of news channels.¹⁰

The evidence from studies conducted in various countries on the effectiveness of fact-checking is mixed. Methodologies range from studies using a quantitative experimental design to qualitative studies involving focus groups, interviews and qualitative surveys.

⁶ Displays a network of organizations with a core/periphery structure linked by joint session participation at Global Fact conferences 2014–2021. This includes dedicated fact-checking organizations, journalism and civil society organizations, technology companies, academic institutes and others (Lauer & Graves, 2024).

⁷ Walter *et al.* (2020, p. 351).

⁸ Quelle & Bovet (2024).

⁹ Procter *et al.* (2023), supported by the Engineering and Physical Sciences Research Council (EPSRC), UK and UK Research and Innovation (UKRI).

¹⁰ Kyriakidou *et al.* (2023); qualitative surveys, N = 1065 and N = 542.

- A study involving a total of 28 experiments in which participants in *Argentina, Nigeria, South Africa* and the *United Kingdom* evaluated 22 fact-checks on topics including Covid-19, local politics and crime found that fact-checks reduced beliefs in mis- and disinformation, but there was also evidence of the so-called ‘backfire’ effect (see below).¹¹
- A literature review of empirical evidence from *Brazil, Malaysia, Turkey, the United Kingdom* and the *United States* of the efficacy of fact-checking for combating mis- and disinformation concluded that fact-checking, originally conceived as a way to hold politicians to account, was not able to meet the expectations of policy makers.¹²

The transparency of sources, funds and methodology is argued to be key to engendering public trust in fact-checking.¹³ Fact-checking transparency was compared in another study with a sample of fact-checkers in *Australia, Brazil, India, South Africa, the United Kingdom* and the *United States*.¹⁴ In South Africa it was found that fact-checking undertaken by NGOs scored higher in transparency than fact-checking undertaken by newsrooms and fact-checkers; India scored higher than those in the United States and the United Kingdom. Others have concluded that more work is needed to understand the influence of different political and national contexts on how publics understand and engage with fact-checking.¹⁵

An experiment-based study conducted in *Australia* to investigate public trust in third-party fact-checkers of political claims found, contrary to expectations, that third-party fact-checking had a negative influence on trust in the fact-checked news story – that is, it had ‘a ‘backfire’ effect on strengthening media trust.¹⁶ The results also showed, however, that the public’s trust, in both

news stories and the fact-checks done on them, was dependent on their trust in the media. It was concluded that if third-party fact-checks were intended to help restore trust in media, then they needed to be conducted in different ways. The results of this research may not be generalizable to different countries, and whether they would generalize to claims other than those made by politicians should be considered. In addition, even if a fact-checking process discloses shortcomings that would justify a decline in trust in a given story, it may or may not reduce trust in the source itself, making assessments in this area difficult to interpret.

During elections, fact-checking strategies include detecting, tracking and countering mis- and disinformation and content curation, emphasizing community standards and algorithm changes, with some efforts to demonetize content by requiring platforms to stop those who generate mis- and disinformation from receiving an income from their ‘clickbait’ or ‘counterfeit news sites’. These efforts are complemented by an emphasis on ethical practice, media and information literacy initiatives, empowering individuals and journalists to track mis- and disinformation via content verification tools and web content indicators.¹⁷

A meta-analysis of studies examining the effectiveness of fact-checking in correcting political mis- and disinformation indicates that it can have a positive and significant effect by correcting beliefs, and it can positively affect beliefs ‘irrespective of political ideology, pre-existing positions, context (campaign vs. routine) and whether it refutes the entire false statement or just parts of a statement’.¹⁸ The effects of fact-checking on beliefs are not very strong, however, with research suggesting that effects become almost negligible when study designs mimic real-life scenarios. Fact-checking is more effective for strengthening preexisting beliefs than for countering mis- and disinformation.

¹¹ Porter & Wood (2021), supported by Luminate and the John S. and James L. Knight Foundation and with participation of Full Fact, Africa Check, Chequeado and others.

¹² Vinhas & Bastos (2022), supported by Twitter, Inc.

¹³ Brandtzaeg *et al.* (2018), supported by the European Commission.

¹⁴ Ye (2023).

¹⁵ Kyriakidou *et al.* (2023).

¹⁶ Carson & Gibbons (2023), N = 1608, supported by Meta with no role in the design or findings.

¹⁷ Cipers *et al.* (2023, p. 7).

¹⁸ Walter *et al.* (2020, p. 366).

In the *United States*, studies suggest that perceptions of the objectivity of fact-checking organizations are influenced by political and ideological divides, which reduces the perceived trustworthiness of fact-checks. At the very moment that fact-checking is most needed, that is, during electoral campaigns, audiences may give little attention to fact-checks that contradict their political views and preferences.¹⁹ Fact-checking is, however, important as part of broader efforts to hold power to account. It is also important to recognize that fact-checking can leave a trail of accountability of politicians, which can reverberate in the future by tarnishing their reputations when they promote divisive issues in misleading ways.

Another study during the 2020 United States midterm elections assessed the effectiveness of different interventions to combat mis- and disinformation on X.²⁰ The conclusion was that fact-checking by professional fact-checkers was less effective than faster algorithmic fact-checking, even though the latter was more prone to error. Advances in AI systems, for example, generative AI (GenAI) tools in the form of large language models (LLMs), have succeeded in lowering the costs of mis- and disinformation production, while making it accessible to more people with little technical knowhow or training.²¹ This is typically perceived as increasing the threat of mis- and disinformation. However, a study that examined whether LLMs might also be used to counter these kinds of information demonstrated that, when used with prompts specifically designed for the purpose, LLMs *may be a useful tool* for detecting LLM-generated mis- and disinformation,²² indicating that the applications of these models depend on the motivations of their developers.

It is also important to note that fact-checking is not a static process whose effectiveness can be assessed without reference to changes

in techniques and practices over time or with changes in how people engage with fact-checked information. At the same time, there is an AI systems 'arms race' to make mis- and disinformation less detectable.²³

2.2 DIGITAL PLATFORM STRATEGIES TO COMBAT MIS- AND DISINFORMATION

The policies and practices reported as being adopted by platforms for content moderation include:²⁴

1. Flagging and review of content
2. Filtering, limiting, blocking or removal of content
3. Promotion/demotion of content
4. Disabling or removal of accounts
5. Transparency in sponsored content
6. User involvement
7. Appeal mechanisms

Some of these measures are aimed at addressing declining trust in the news by, for example, labeling news outlets that adhere to international journalism standards, providing tools that can be used to prioritize trustworthy content and promoting trustworthy news to audiences and advertisers.²⁵

Digital platforms are responding to mounting public scrutiny by implementing some measures for countering mis- and disinformation – some temporary and linked to circumstances such as Covid-19 or to specific electoral campaigns. They are reluctant to censor content or permanently ban prominent users (i.e., political figures) who promote mis- or disinformation, although they sometimes do so when they become embroiled in politically partisan disputes. Although they remain focused on using their algorithms to amplify engagement with content (including mis- and disinformation), when they do seek to moderate what they determine to be unacceptable content, their preferred approach

¹⁹ Walter *et al.* (2020).

²⁰ Guriev *et al.* (2023), funded by the Project Liberty Institute, US

²¹ Xu (2023).

²² Jiang *et al.* (2023).

²³ Innes *et al.* (2023).

²⁴ Bontcheva *et al.* (2020), Table 5 'Curatorial responses from internet communication companies', pp. 145–146 for a list of content moderation policies by platform.

²⁵ Mazzoli (2023); these include NewsGuard (a US non-profit supported by the Knight Foundation and Publicis Groupe, France), The Trust Project (a US non-profit funded by Craig Newmark Philanthropies, Google, Democracy Fund, John S. and James L. Knight Foundation, Facebook and the Markkula Foundation) and the Journalism Trust Initiative (Reporters sans frontières [Reporters Without Borders]), which focuses on technical standards for journalists and media outlets.

is to act against suspicious behavior on user accounts by deprioritizing the visibility of mis- and disinformation in news feeds using a mix of automated and human curation content moderation practices.²⁶ Furthermore, it is argued that when platforms – in this case, Facebook, YouTube and X – aim to promote their own growth, the views of communities of color, women, religious minorities and LGBTQ+ people are ignored, and the harms associated with content targeting them remain unaddressed.²⁷

The role of digital platforms in sustaining trustworthy information needs to be monitored since they are known to roll back measures to combat mis- and disinformation, lay off staff or weaken privacy policies and impose limits on user fact-checking.²⁸ They have done so at the same time as they signed a voluntary AI Elections Accord in 2024 to take ‘reasonable precautions’ to prevent AI tools from being used to generate content that contributes to the disruption of elections as a result of the circulation of inaccurate or false information.²⁹

The major digital platform owners (Meta, Google [YouTube] and X) are criticized for engaging in a new form of ‘private multilateralism’ in their efforts to comply with mis- and disinformation governance, going beyond requirements in an effort to mitigate reputational damage or to ward off legislation, granting the platform owners too much discretion over who is silenced.³⁰ An example of silencing is when a platform bans a user’s content from appearing without notifying the user, with the clear effect of stifling the user’s speech. This is a type of harm that receives less research attention in comparison to research on the harms associated with the virality of content.³¹ The platforms’ content detection and takedown tools are also criticized for being designed to tackle very specific forms of

mis- and disinformation or propaganda, for example ISIS- or al Qaeda-generated content, with the result that the tools do not work well in other contexts.

It needs to be acknowledged that ‘there is no such thing as a “neutral” recommendation algorithm’, and it is crucial to avoid simple solutions such as insisting on ‘non-discrimination’.³² Suggestions for mitigating the biases of the platforms’ algorithmic personalization systems include offering users different ranking options in their feeds as a means of enhancing their agency, in addition to supporting media and information literacy programs and boosting public transparency around the design and performance of personalization systems.³³ As far as the design of platforms is concerned, there are suggestions that various forms of content moderation do not address the need to provide disincentives for the production of mis- and disinformation. These might include changes in content ranking so as to reward productive interactions, limits on the dissemination of content on platforms, and alterations to the design features of platforms that ‘support the on-platform work of peacebuilders’, to facilitate information ecosystems in which everyone is able to thrive.³⁴

2.3 EFFECTIVENESS OF MIS- AND DISINFORMATION COUNTERMEASURES

Diverse governance approaches by states to content moderation may be justified by the fact that the platforms offer different kinds of services, but this means that there is no consistency in terms of enforcement, transparency or appeal against state or platform content moderation decisions.

The European Union’s Code of Practice on Disinformation model of content governance has been taken up beyond the region, and most

²⁶ Saurwein & Spencer-Smith (2020).

²⁷ Díaz & Hecht-Felilla (2021).

²⁸ Kennis (2024).

²⁹ O’Brien & Swenson (2024). For an update on commitments made by signatories (Adobe, Anthropic, ElevenLabs, GitHub, Google, LG AI Research, LinkedIn, McAfee, Meta, Microsoft, Nota, OpenAI, TikTok and Truepic) to September 2024, see AI Elections Accord (2024).

³⁰ Borelli (2023).

³¹ Horten (2023).

³² Llansó *et al.* (2020, p. 3).

³³ Literacy is discussed in Chapter 5, as is public awareness.

³⁴ Stray *et al.* (2023, p. 6).

platforms are employing some form of content moderation or oversight system, but within the European Union, the transparency of content removal policies and practices is criticized in some research for being weak.³⁵ Code implementation monitoring indicates its usefulness as a tool, but reveals room for improvement: the need for clearer procedures, more precise and comprehensive commitments, the need for transparent key performance indicators (KPIs) and for appropriate monitoring.³⁶ The measures are criticized for being too general in terms of content and structure; for lack of compliance and enforcement tools; the absence of a shared terminology (impacting on the focus of actions under the provisions of the Code); and a lack of granular platform data needed for independent assessment. Also mentioned are: the risk of de facto privatization of censorship (i.e., delegation of important content-based choices to platforms); insufficient collaboration; unclear or unstated reporting on regional actions; and a lack of standardized procedures to verify the implementation of actions under the Code.³⁷

Despite these criticisms, a study examining how global digital platforms have responded to this Code of Practice, based on their annual reports, concludes that as a form of 'soft governance', the Code can be effective under certain conditions, due to flexible reporting and the higher priority given to certain issues by the digital platforms. It is also observed, however, that 'soft governance permits companies to window dress low compliance by reporting cost effective operational changes'.³⁸

Investment in these mechanisms varies substantially across regions and languages. It is concentrated in the Global North, which also sees more consistent

enforcement of the platforms' own rules. Smaller markets and content in languages other than English receive much less attention.³⁹ Systems for automatic content moderation are largely trained on English language content because of the gap between the quantity, quality and diversity of training data in English versus other languages.⁴⁰

The International Panel on the Information Environment (IPIE) reported on the effectiveness of a broad range of countermeasures designed to tackle mis- and disinformation, including, but not limited to, legislation. It was found, in 2023, that fewer than one-fifth of the studies surveyed reported that countermeasures were effective, and that effectiveness depended on the type of intervention and types of information, with uncertainty about the size of the effects.⁴¹

When the big tech companies signed their AI Elections Accord they acknowledged that 'deceptive AI Election content can deceive the public in ways that jeopardize the integrity of electoral processes'. They called for a 'whole of society approach', and said that no individual solution, or even combination of solutions, would 'fully mitigate the risks'. They also called for education of the public to mitigate risks.⁴²

Mis- and disinformation campaigns during elections expose shortcomings in electoral regulations, especially in relation to the use and exploitation of social media platforms. The ability of people to participate in the public sphere is strongly influenced by the effectiveness of co-regulation between governments, electoral authorities and the digital platforms.

³⁵ Solomun *et al.* (2021); see also EC (2022e) and Section 4.2, Chapter 6.

³⁶ EC (2020a).

³⁷ ERGA (2020); Monti (2020); Nenadić *et al.* (2023); Pamment (2020).

³⁸ Borz *et al.* (2024, p. 18).

³⁹ Dergacheva *et al.* (2023); Global Witness (2023); UN (2023a).

⁴⁰ Nicholas & Bhatia (2023).

⁴¹ See IPIE (2023a, b, c), based on a systematic review of 588 peer-reviewed studies of the effectiveness of countermeasures, an expert survey on trends in the global information environment and platform responses.

⁴² Adobe *et al.* (2024, p. 1).

False narratives in the public sphere.

A study of institutional and regulatory frameworks in 2024 showed the absence of effective policies for the fight against mis- and disinformation. For example, 46% of women and girls in Central and West Africa who participated in the study thought that false narratives were being used to dissuade women from participating in the public sphere.⁴³

For some time now some governments have had laws in place against the communication of specific types of false communication during elections. Since these were not tailored to scenarios where actors are taking advantage of the affordances of digital platforms to rapidly and covertly disseminate misleading information, countries are updating their regulations, notably *France* and *Australia* (whose new regulations focus on foreign interference in elections). Other measures are being directed towards the transparency of political online advertising, one approach being the creation of publicly accessible ad archives. Many countries and regions have proposed or passed new laws in this area (e.g., *Brazil*, *Canada*, the *European Union* and member states – *France* and the *Netherlands* – *India*, the *United Kingdom* and the *United States*).⁴⁴

Brazilian banning order on X. In August 2024, a Brazilian court judge issued a banning order on X after Elon Musk refused an earlier order to suspend more than 140 accounts, which the same judge had determined were a threat to Brazilian democracy. The judge declared that Musk was allowing ‘the massive spread of disinformation, hate speech and attacks on the

democratic rule of law, violating the free choice of the electorate, by keeping voters away from real and accurate information’.⁴⁵ Musk responded that the judge was ‘an evil dictator’.⁴⁶ It was subsequently reported that X had circumvented the ban by an update to its communications network, which enabled alternative routes for Brazilian users to access X.⁴⁷

This episode is a classic case of tension between freedom of speech and freedom from harm. It is also indicative of struggles by countries to introduce policies that respect national sovereignty and the rule of domestic law.⁴⁸ In other cases, X has complied with government demands to ban accounts and content deemed to be critical as, for example, when X demoted or banned content critical of India’s Prime Minister Narendra Modi during the farmers’ agricultural policy protests in 2024.⁴⁹

The impetus to counter mis- and disinformation can be weaponized by authoritarian governments to suppress opposition and to control the flows of information.⁵⁰ Risks associated with anti-disinformation policies are high when policies grant excessive power to government bodies, ruling political parties or other established actors, enabling them to manipulate communication during critical election periods or other crises.⁵¹ Digital technologies are being used by governments as tools for surveillance, control and censorship, which deters political expression and engagement. There is also evidence of the increasing involvement of the military, police, state information services and security consultants, which presents a potential or real threat to democracy.⁵²

⁴³ Zibi Fama (2024).

⁴⁴ Tenove (2020).

⁴⁵ Nicas & Conger (2024).

⁴⁶ Shapero (2024).

⁴⁷ Murphy *et al.* (2024).

⁴⁸ Curzi de Mendonça (2024).

⁴⁹ Sankaran (2024).

⁵⁰ See Section 4.3.3., Chapter 2 for evidence from different regions.

⁵¹ Tenove (2020).

⁵² Casero-Ripollés *et al.* (2023).

Of 11 laws introduced between 2016 and 2020 to respond to ‘false information’ in *Benin, Burkina Faso, Côte d’Ivoire, Ethiopia, Kenya, Malawi, Niger, Nigeria, South Africa* and *Uganda*, for example, these 10 countries were found to have passed or amended laws or regulations penalizing the publication or broadcast of what the authorities considered to be false. The 11th, *Senegal*, used existing laws for the same purpose.⁵³ Given the arbitrary definitions of false content that have been adopted, such laws are found to have an unwelcome effect on freedom of expression and to fail to limit the spread of mis- and disinformation.⁵⁴

Motivation to share mis- or disinformation.

In evaluating the effectiveness of governance responses to mis- and disinformation it is important to recognize that in some countries, for example African countries, politically motivated mis- or disinformation, often expressed as satire, was aimed at ridiculing elites and was common before this became a preoccupation in the Global North. The spread of mobile phone use exacerbated the problem of mis- and disinformation, but scholars emphasize that analysis needs to take values into account, including solidarity, interconnectedness and interdependence, which help to explain why people are motivated to share mis- or disinformation.⁵⁵

Democracies with high levels of press freedom are more likely to take a more holistic approach to countering online disinformation, focusing comparatively more on the integrity of their election process, media and education initiatives.⁵⁶ Countries with a higher GDP have more initiatives and legislation in place than countries with a lower GDP. A study analyzing government responses to online mis- and disinformation in 103 countries, and 10 international and regional organizations

across six continents, found substantial differences between democratic and authoritarian regimes. Authoritarian regimes seemed to prefer responses that were sufficiently vague to allow them to repress opposition and criticism. Examples include a cybercrime law in *Zimbabwe*, enforcement of Article 80 of the Constitution in *Tunisia*, and the cybercrime laws in *Nicaragua*. The less democratic the country, the less likely it is to implement election-specific responses to maintain the integrity of the electoral process. And these repressive acts are not limited to authoritarian states (such as *China, Iran* and *Russia*); they also occur in democracies such as *India* and *Mexico*.⁵⁷

Problems in combating mis- and disinformation are exacerbated by the economic fragility of news media organizations, which makes them susceptible to capture by governments. More legal safeguards are clearly needed to prevent the risks of political influence.⁵⁸ Countermeasures to mis- and disinformation can have a negative impact on freedom of expression, and even attempts in good faith to govern in response to mis- and disinformation can be controversial. For example, in 2018 *Italy* enacted the ‘Operating Protocol for the Fight Against the Diffusion of Fake News through the Web on the Occasion of the Election Campaign for the 2018 Political Elections’, giving the Postal and Communications Police (*Polizia Postale e delle Comunicazioni*) authority to determine whether online claims were false or biased, and to recommend judicial action. This was widely criticized by experts, the United Nations Special Rapporteur for the protection of freedom of expression and journalist organizations.⁵⁹

Proposals to assist in countering mis- and disinformation insist on ‘meaningful transparency’ in the way platforms operate, especially regarding their content moderation practices, and specifically by mandating platform ‘interoperable transparency’. Developed by the Internet Governance Forum, the

⁵³ Cunliffe-Jones (2021).

⁵⁴ Cunliffe-Jones (2021).

⁵⁵ Morales *et al.* (2021).

⁵⁶ Cipers *et al.* (2023).

⁵⁷ Cipers *et al.* (2023), 239 responses comparing level of democratization, press freedom and GDP with types of responses.

⁵⁸ Radsch (2023d).

⁵⁹ Kaye (2018).

requirement is that reports published by platforms should provide both quantitative and qualitative contextual information when they make data available for research purposes. They should share detailed and intelligible information in standardized formats, and make data continuously available in an interoperable, understandable and machine-readable format that is auditable by third parties.⁶⁰ In this way it might become feasible to provide clearer explanations of how platform operations exacerbate the problem of mis- and disinformation.

3 The Challenges of Information Ecosystems Governance

This section examines some of the main challenges of information ecosystems governance – efforts to improve the accountability and transparency of AI systems, to foster news media independence and press freedom, and govern data in corporatized information ecosystems.

3.1 GOVERNING AI SYSTEMS: ACCOUNTABILITY AND TRANSPARENCY

When AI systems are deployed for content governance, it presents challenges and opportunities for the sustainable design of information ecosystems. AI technologies remain, to a certain degree, unregulated, even though there are more regulations. In September 2024, the United Nations AI Advisory Board's *Governing AI for Humanity: Final Report* stated that voluntary commitments to mitigate the harmful effects of AI generated content 'that is likely to deceive in a way that causes harm or societal divisions, or which

promotes war propaganda, conflict and hate speech ... *do not sufficiently mitigate societal risks*. Instead, a global, multi-stakeholder approach is required, along-side binding commitments.⁶¹

Approaches in the United States and Europe, including the European Union's AI Act, encompass the prohibition of certain kinds of AI applications, stringent rules for high-risk systems (such as certification schemes) and broad transparency obligations that mirror concerns about the 'black box' nature of advanced ML systems. As these systems are always used in specific social contexts, additional regulation within these contexts also usually applies.⁶² Many of these provisions have a bearing on the production and circulation of mis- and disinformation.

To effectively regulate AI systems, it is vital to understand how the technology is being used, and how this impacts on democratic decision-making systems and on protected rights and values. Research emphasizes that international standards and governance approaches, regional norms and national laws need to be intertwined to effectively meet the challenges of automated content governance systems.⁶³ However, the 'soft law' character of many existing rules (such as commitments to limit mis- and disinformation), apart from the European Union's AI Act (which had yet to enter completely into force at the time of writing), makes assessing regulatory effectiveness very difficult.

The AI Act includes specific mention of systemic risks through 'the facilitation of disinformation or harming privacy with threats to democratic values and human rights'.⁶⁴ It also states:

Furthermore, obligations placed on providers and deployers of certain AI systems in this Regulation to enable the detection and disclosure that the outputs of those systems are artificially generated or manipulated are

⁶⁰ Belli *et al.* (2022).

⁶¹ UN (2024a, p. 35, 36, emphasis added).

⁶² See Chapter 6.

⁶³ For a discussion of standards, see von Ingersleben-Seip (2023).

⁶⁴ EC (2024c, para. 110).

particularly relevant to facilitate the effective implementation of Regulation (EU) 2022/2065. This applies in particular as regards the obligations of providers of very large online platforms or very large online search engines *to identify and mitigate systemic risks that may arise from the dissemination of content that has been artificially generated or manipulated, in particular risk of the actual or foreseeable negative effects on democratic processes, civic discourse and electoral processes, including through disinformation.*⁶⁵

Some critics are not satisfied that policy interventions such as the European Union's Digital Services Act (see Chapter 6) and the AI Act will be effective in countering mis- and disinformation. They argue that regulations on the use of AI systems in the generation, moderation and curation of content streams need to be better informed by evidence of the nature of the threats and effectiveness of mitigation measures.⁶⁶ Others point to the need for effective verification processes for assessing compliance, without which 'this regulation risks being as ineffectual as GDPR, yielding only marginal improvements in ethical AI system practices'.⁶⁷

New AI policies. Since 2016 it is estimated that some 800 AI policy initiatives in at least 60 countries have been discussed, with the aim of tailoring AI governance to specific country conditions in a way that respects human rights and results in transparency and accountability.⁶⁸

Yet others call for clarity on the normative goals of AI systems governance. There needs to be a shared understanding about what the societally appropriate balance between economic and innovation imperatives and human rights should be.⁶⁹ The regulatory challenge is to decide on that balance, and then design legislation that is capable of delivering and sustaining it against the risks, for example, of regulatory overreach by governments, and particularly authoritarian governments.⁷⁰

Approaches will differ from country to country.⁷¹ For AI systems generally, and GenAI specifically, regulatory solutions need to be approached democratically and on a global scale, with distinctive approaches that acknowledge that conditions vary between the Global North and Global Majority World. No set of governing institutions is likely to be regarded as ideal by all parties concerned, and the key issue is to ensure that whatever institutions are in place at the global, regional and national levels, they are used in ways that are consistent with human rights protection and the public's interest.⁷² For the latter, allowing private technology firms and regulators primarily based in the Global North to set the parameters for AI regulation poses significant risks.⁷³ These include the risk that choices about AI systems design and deployment, made mainly in the Global North, will be inappropriate in different cultural, social and political contexts. There is also a risk that the largely hidden ability of big tech companies to exploit relatively cheap labor in the Global Majority World during AI systems development will become more widespread if regulatory approaches do not take the processes of workforce extractivism into account.⁷⁴

⁶⁵ EC (2024c, para 120; emphasis added).

⁶⁶ Nannini *et al.* (2024) supported in part by European Union Horizon program and the Galician Ministry of Culture, Education, Professional Training and University.

⁶⁷ Schmitt *et al.* (2024, p. 36), supported by the Federal Ministry of Education and Research (BMBF, Bundesministerium für Bildung und Forschung), Germany.

⁶⁸ Roberts *et al.* (2023).

⁶⁹ Cornils (2020), supported by AlgorithmWatch, European Policy Centre (EPC), Mainzer Medieninstitut und Civitates

⁷⁰ Marda & Milan (2018).

⁷¹ Ashok (2023), funded by Platform Regulation and Operations in the Sharing Economy Project, Australian Research Council (ARC).

⁷² Souter (2024).

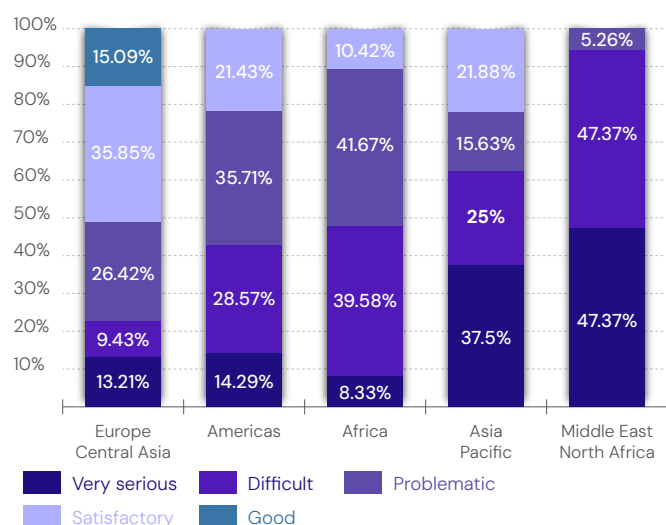
⁷³ Ross Arguedas & Simon (2023).

⁷⁴ Monasterio Astobiza *et al.* (2022).

3.2 GOVERNING NEWS MEDIA AND PRESS FREEDOM

The production and circulation of both accurate information and mis- and disinformation are crucially dependent on the extent to which the news media in a country enjoys press freedom, that is, independence from the state or from corporate owners. The degree of press freedom can influence whether journalists feel secure and safe enough to produce content that can be assessed as fairly reporting on the conditions in a given country. Reporters sans frontières (RSF, Reporters Without Borders) defines press freedom as ‘the ability of journalists as individuals and collectives to select, produce and disseminate news in the public interest independent of political, economic, legal, and social interference and in the absence of threats to their physical and mental safety’.⁷⁵ Measured using indicators of political context, legal framework, economic context, socio-cultural context and safety, Figure 7.2 shows the state of press freedom in 2024, with the World Press Freedom Index showing a global average decline of 7.6 points on the political indicator since 2013. Only 15% of countries achieved the (green) ‘good’ press freedom status.⁷⁶

Figure 7.2
World Press Freedom Index, 2024



Source: RSF (2024).

Between 2016 and 2020, 400 journalists were killed, and in 2020, 73% of 625 surveyed female-identifying journalists experienced some form of online violence.⁷⁷ Governance measures to uphold the right to freedom of expression and freedom from harassment and violent attacks need to tackle structural inequalities, technological change, political alignments and social transformation.⁷⁸ In the case of news media, these measures include efforts to strengthen newsroom practice and financial sustainability, and preventing the silencing of journalists’ voices; for example, the Organization for Security and Co-operation in Europe’s (OSCE) Group of Friends on the Safety of Journalists is calling for the implementation of commitments to protect journalists.⁷⁹ Other actions include the United Nations’ *Plan of Action on the Safety of Journalists and the Issue of Impunity*⁸⁰ and the work of the Freedom Online Coalition.⁸¹ In these initiatives, it is important to acknowledge the subtleties, trade-offs and negotiations that journalists engage in when ‘they try to understand, anticipate, leverage, or resist how social media platforms see and organize news’.⁸² Measures are needed to protect independent journalism – ‘independent journalism needs freedom, funding, and a future’⁸³ – but the evidence confirms that the situation is not improving in many countries.

In the early 2010s, *Germany* and *Spain* legislated in support of the news media in their relation to Google’s Search and News products, with mixed results. *Australia* and *Canada* designed reforms in response to the unequal bargaining power between digital platforms and news media publishers (the News Media Bargaining Code and the Online News Act, respectively). Other counties may introduce similar legislation, for example, *India*, *New Zealand*, the *United Kingdom* and the *United States*. Skeptics argue, however, that these measures are centered around disputed claims about the economic

⁷⁵ RSF (2022).

⁷⁶ RSF (2024).

⁷⁷ UNESCO (2022d, pp. 17, 20).

⁷⁸ Davis *et al.* (2020).

⁷⁹ Kamp (2023).

⁸⁰ UN (2023c).

⁸¹ Freedom Online Coalition (2023).

⁸² Ananny (2023, p. 1570).

value of news, and that they will not recapture the advertising revenue lost to digital platforms. This means that the long-term relationship between the news media and platforms remains contested.⁸⁴ These developments can weaken the efforts of news media organizations to produce accurate news and to provide counters to mis- and disinformation.

UNESCO notes that without sufficient funding, news media ‘independence can be easily compromised and pluralism becomes a shadow of what it should be’.⁸⁵ Overseas development aid can help to sustain journalism and independent media in the Global Majority World when it provides scope for local initiatives and control over changes. From 2010 to 2019, however, aid in support of initiatives worldwide to sustain the media stagnated as a proportion of total aid, at 0.3% of USD 200 billion, or an average of USD 300 to USD 400 million annually, for initiatives worldwide to sustain the media.⁸⁶

In addition to the big tech platforms, intermediary companies operate, claiming to free news media organizations from dependency on the largest platforms. A study of platforms self-describing as ‘content recommender platforms’ and claiming to free up news organizations in this way – Taboola and Outbrain – indicates that although they operate as ‘match makers’ between advertisers, publishers and data brokers, they blur news content with sponsored content, making it difficult for audiences to detect the difference.⁸⁷ Even when smaller companies play a role in making news available, they operate according to economic incentives that lack transparency and depend on corporate decisions about news priorities. And, research on strengthening the contributions of the news media industry to a healthy information ecosystem is heavily focused on the United States and Europe.

There are many studies of measures to sustain independent, high-quality trustworthy news in the face of the growth of mis- and disinformation and globally declining newspaper sales, for example, tax credits, direct subsidies, philanthropic funding, not-for-profit models and subscription or membership models. However, none of these on their own are likely to fully address the news media industry’s funding gaps or strengthen its ability to combat mis- and disinformation.⁸⁸ One challenge is demonstrating that increasing a public subsidy or introducing a tax is substantial enough to compensate for the historical cross-subsidy of news content by advertisers to organizations funded in this way. With advertisers now attracted to the platforms rather than directly to news organizations, it is argued that the cost burden falls ultimately on the news-consuming audience, and it must be demonstrated that news has value for the health of the public sphere and democracy beyond the measures used to value news economically.

3.3 GOVERNING DATA IN CORPORATIZED INFORMATION ECOSYSTEMS

Efforts to combat harmful mis- and disinformation must address the governance of the data – the lifeblood of online systems that enable content creation and circulation. Data is generated through online engagement. This data is the input to algorithmic personalization systems that determine who will encounter different kinds of online mis- and disinformation.

Strong data and privacy protection regulatory frameworks emphasizing compliance and promoting international cooperation aim to create a secure and rights-respecting data collection and processing environment. These are crucial for maintaining public trust in platform services, AI

⁸³ Nielsen *et al.* (2019, p. 11), in Brazil, India, the United Kingdom and the United States.

⁸⁴ Flew (2023), supported by the Australian Research Council (ARC); Flew *et al.* (2023); Flew & Stepnik (2024), supported by Australian Research Council (ARC).

⁸⁵ UNESCO (2022, p. 7).

⁸⁶ Myers & Gilberds (2024), based on interviews with agency representatives, supported by the National Endowment for Democracy (NED), a quasi-autonomous NGO in Washington DC, US.

⁸⁷ These operate in Israel and embed their widgets in news media websites such as the BBC, CNN or The Washington Post, NBC News and FOX News; see Ratner *et al.* (2023), supported by the European Research Council (ERC).

⁸⁸ See Borges & Reviglio (2024); Brogi & Sjøvaag (2023); Henningsen & Krčál (2024); Kennis (2024); Mansell & Steinmueller (2020, pp. 89–94); Mutsvairo (2019); Pickard (2020c, 2024); Rohn *et al.* (2024).

systems and the news media. They are intended to enable the beneficial aspects of data collection to be realized, but they must do so in a context where the monetization of data for profit by big and smaller companies is promoted in the name of economic growth, while also minimizing potential harms and infringements of human rights.

For this reason, the aim of state governance interventions is to ensure that data-driven products, especially those employing AI, involve ethical collection and use of data to ensure transparency about how data is used. Ideally, these governance measures should mitigate algorithmic bias and prevent discriminatory outcomes. Although privacy and data protection legislation is intended to prevent harmful outcomes,⁸⁹ the evidence indicates that people's rights are inadequately protected, despite legislative measures.

One reason is the global nature of cross-border data transfer operations. This presents challenges in ensuring that data transfers comply with robust privacy and data protection standards. These operations involve jurisdictional complexities when data handled overseas must be accorded the same level of protection as within the originating country.⁹⁰ This is especially difficult when digital platforms weaken or compromise their privacy and data protection practices in line with their commercial interests and competitive pressures to enhance user engagement and advertising revenue.⁹¹

In most countries, the collection and use of specific types of data by digital service providers — such as biometric data, health information and personal identifiers — are subject to stringent governance due to their sensitive nature.⁹² To be effective, a complex legal and ethical landscape must be navigated. Studies show that the mere existence of data protection laws does not

translate automatically into effective enforcement. In addition, the availability of adequate resources, financial and human, is crucial for data protection authorities to perform their functions, and this varies by country.⁹³

Examples of claims to effective privacy and data protection compliance. In the United-States, for example, the Health Insurance Portability and Accountability Act (HIPAA) is claimed to be well-enforced, with significant penalties for non-compliance given the high level of sensitivity of health data. Companies that specialize in data analytics and involvement with healthcare data navigate the HIPAA by leveraging their technological expertise to comply, while capitalizing on the commercial value of the data. Palantir claims that its data management system and its advanced security protocols enable it to handle sensitive health information efficiently and securely, thereby meeting the requirements of the HIPAA.⁹⁴

This example illustrates how the data economy comes to be presented as beneficial for all. The commercial success of data analytics companies in the health and other sectors shows how organizations with substantial resources can thrive under stringent legislation by integrating compliance into their operational models and using technology to manage compliance requirements. Small healthcare operators claim that they are burdened by the costs of compliance with data protection requirements,⁹⁵ and there is a robust literature on how companies like Palantir are using data in ways that discriminate among different segments of the population.⁹⁶ There is growing experience with data anonymization algorithms, but there is always a risk

⁸⁹ Mitchell *et al.* (2021).

⁹⁰ Reis *et al.* (2024).

⁹¹ Floridi (2013).

⁹² Payne *et al.* (2023).

⁹³ Bennett & Raab (2006).

⁹⁴ Palantir specializes in software platforms for big data analytics, see for compliance statements <https://www.palantir.com/attributes/security/>.

⁹⁵ Chen and Benusa (2017). Calls by leading companies in the development of LLMs for regulation have similarly met with suspicion that their real goal is to stifle competition.

⁹⁶ Ulbricht & Egbert (2024), supported by Bundesministerium für Bildung und Forschung (Weizenbaum Institute for the Networked Society) and the European Research Council (ERC).

of re-identification, and anonymity can be reversed. In addition, it is unclear how these methods compare, and which are most suitable for achieving anonymization for specific purposes.⁹⁷

At the same time, there are studies of how Palantir's expansion into the health sector in the United States brings the risk of many transgressions beyond issues of data protection, as the company engages in developing invasive surveillance software technology, for both the health sector and its security business.⁹⁸ And in countries beyond the United States and across sectors there are concerns about how robust privacy and data protection legislation is. Examples of areas in which the potential for human rights abuses of data are of major concern include the use of systemically biased data in facial recognition systems used in efforts to identify individuals participating in crime, including riots or, for example, to identify migrants' children, another area where Palantir is involved especially in predictive policing.⁹⁹

It is also argued that policy makers tend to focus myopically on content and platform governance issues and data protection, ignoring the internet itself as a key tool for the spreading of mis- and disinformation. Greater attention to internet intermediaries, including Domain Name System (DNS) registries, financial intermediaries and Internet certification authorities, is needed if the harms of mis- and disinformation are to be tackled effectively.¹⁰⁰

4 Human Rights and Mis- and Disinformation Countermeasures

The United Nations General Assembly comment on mis- and disinformation emphasizes the need for governance measures to combat mis- and disinformation while respecting human rights. It cautions that 'simple solutions to this complex problem are likely to censor legitimate speech that is protected under international human rights law'.¹⁰¹ It also warns against the use of vague definitions of mis- and disinformation and strategies deploying excessive and disproportionate sanctions, and calls for involving diverse stakeholders in developing governance strategies. The interim report by the United Nations Advisory Board on AI highlights the challenge: while new methods of mis- and disinformation represent a real threat to political processes, strategies to combat mis- and disinformation risk infringing human rights.¹⁰²

Evidence on efforts to implement mis- and disinformation governance measures is mixed on whether these measures adhere to the human rights principles that countries are signed up to.¹⁰³ Based on a review of more than 800 sources between 2020 to 2022, a UNESCO report found that, while many countries have laws in place governing online content, they are often vague, leading to arbitrary measures by public authorities and corporate platform operators. There is a gap between generic policy and how local issues are dealt with, especially when rights protection and business model goals are in tension.¹⁰⁴ Even though large platform companies have pledged support for the United Nations Guiding Principles on Business and

⁹⁷ Sepas *et al.* (2022).

⁹⁸ Lanzing (2023).

⁹⁹ Chouliraki & Georgiou (2022); Lanzing (2023).

¹⁰⁰ Bradshaw & DeNardis (2022).

¹⁰¹ UN (2022, p. 11).

¹⁰² UN (2023b).

¹⁰³ Rajkumar & Ashraf (2023).

¹⁰⁴ Berger *et al.* (2023).

Human Rights,¹⁰⁵ they often fall short. In addition, when multilevel governance measures are in place (as in the European Union), national strategies and practices can provide uneven protections, and there are struggles to coordinate the multiple initiatives.¹⁰⁶

Task forces, expert groups and think tanks are proliferating with the aim of influencing policy and practice. In *sub-Saharan Africa* in 2022, of 46 countries, 20 pieces of legislation against ‘disinformation’ had passed under penal codes, two countries had legislated against falsehoods, nine countries had addressed ‘disinformation’, focusing on the intent to mislead or deliberately spread false information, and eight countries were specifically penalizing ‘misinformation’.¹⁰⁷ In *Asia*, critics of measures to combat mis- or disinformation argue that these are used as ‘a cover to silence dissent and legitimate criticism of the state’.¹⁰⁸ In many countries, legislation to address mis- or disinformation draws on laws dealing with defamation, sedition, technology regulation, cybercrime and online harassment, which have been in place since colonial times, and have potentially detrimental implications for the right to freedom of expression.

In addition, a study in 2024 of 32 countries (most in the Global Majority World) that have introduced legislation on mis- and disinformation observed that definitions of mis- and disinformation, and penalties for distributing it, vary substantially among countries. Generally, there is an absence of a clearly designated authority to determine what constitutes the information targeted by legislation. The authors concluded that much legislation ‘*risks doing more harm than good*’, in part because 14 of 32 policies give the government authority to decide what is ‘fake news’.¹⁰⁹

Online activity can support protests against authoritarian regimes just as it can lead to surveillance when people generate trace data that bolsters the interests of authoritarian states and enables commercial datafication practices.¹¹⁰ Campaigns against mis- and disinformation by civil society organizations can also support causes that are pro-government or attack opposition political leaders. When the ambition is to distract, suppress or polarize opinion,¹¹¹ the affordances of datafied information ecosystems can become pro- or anti-democratic, and governance initiatives in the political sphere may favor legislation that does not uphold or defend human rights.¹¹²

In the case of the news media, research on data quality assessment frameworks for AI systems used in journalists’ workflows emphasizes the accuracy, completeness and consistency of data quality from the data acquisition stage onwards, the need for rules, routines and institutional practices to be in place, and improved data literacy.¹¹³ The importance of assigning responsibility to someone in the news organization, experimenting with AI-enabled practices with human oversight, and collaborating with AI systems developers, is highlighted as especially important. Clear lines of accountability for the impacts of automated decisions are needed to ensure that responsible parties can be held accountable for the consequences of algorithm-based actions, and to ensure that people’s rights to freedom of expression, privacy and autonomy are protected.¹¹⁴

Research indicates that government-mandated shutdowns of the internet, and other measures that impact on online expression during elections, is not seen by citizens as being done in the public interest, and it is argued that the priority in response to vicious circles of

¹⁰⁵ Ruggie (2011) for report of Special Representative of the United Nations Secretary-General for the principles endorsed by the Human Rights Council and an explanation.

¹⁰⁶ Saurwein & Spencer-Smith (2020), funded by the Austrian Academy of Sciences (ÖAW, Österreichische Akademie der Wissenschaften; Hatfield *et al.* (2023).

¹⁰⁷ Sey *et al.* (2022, p. 159).

¹⁰⁸ Jayasinghe *et al.* (2022, p. 192).

¹⁰⁹ CNTI (2024, p. 1, emphasis added); 12 designated as democracies, 19 as authoritarian, 1 unclassified, supported by Craig Newmark Philanthropies, the John D. and Catherine T. MacArthur Foundation, John S. and James L. Knight Foundation, Lenfest Institute for Journalism and Google.

¹¹⁰ Mutsaers & Rønning (2020).

¹¹¹ Bradshaw & Howard (2019); Richter (2024).

¹¹² Bhatia *et al.* (2023); Borelli (2023).

¹¹³ Dierckx *et al.* (2023b), supported by the European Commission.

¹¹⁴ Beckett & Yaseen (2023), supported in part by Google; see also Alufaisan *et al.* (2021).

mis- and disinformation should be to restore faith in journalism, not to censor digital media.¹¹⁵

It is crucial to distinguish between mis- and disinformation governance measures that support human rights and those that enable 'privatized censorship' by digital platforms, government censorship with technical measures that violate privacy and blanket indiscriminate censorship.¹¹⁶ As one analyst put it: 'regulation imagines itself as simply enforcing the given and natural norms of a decent society...; but from a sociological or anthropological point of view we know that law is always actually enforcing the mores of the dominant group that controls the content of law'.¹¹⁷

In addition to mis- and disinformation strategies being plagued by the challenge of defining what counts as mis- or disinformation, government measures are criticized for being disproportionate or for only tinkering with mis- and disinformation problems in some jurisdictions.

Criticisms of government strategies. In *Brazil*, the government's policy of leaving it to platform companies to moderate election-related content was found to have resulted in opaque criteria to detect inappropriate content, with platform decisions being likely to have an 'advertising effect', endorsing specific candidates.¹¹⁸

In *India*, the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (IT Rules), updated in 2023, regulate digital media platforms through the Ministry of Electronics and Information Technology (MEITY). These apply to social media intermediaries, digital news publishers and digital platforms. Immediately faced with a severe backlash from civil society and digital

rights organizations, the implementation of parts relating to fact-checking were put into abeyance by the Supreme Court.¹¹⁹

In *Nigeria*, government strategies are seen as threatening to close down the media space through the 2019 Protection from Internet Falsehoods and Manipulation and Other Related Matters Bill, which criminalized the sharing of information that diminishes public confidence in the government. This was seen as silencing the voices of people who are critical of the government.¹²⁰

Policy makers tend to focus myopically on platform governance when they tackle mis- and disinformation, ignoring the internet itself as a key tool for spreading this information. Greater attention to internet intermediaries, including DNS registries, financial intermediaries and internet certification authorities, is needed if the harms of mis- and disinformation are to be tackled effectively.¹²¹

5 Public Appetite for Combating Mis- and Disinformation

Research finds that people's awareness of and ability to navigate information ecosystems varies by country, gender, race, class, culture, socio-economic position, and more. It is not surprising, therefore, that when questions are asked about what the public would like states, the private sector or other organizations to do to mitigate harms linked to mis- and disinformation, the answers depend on the experience of different aspects of these ecosystems.

¹¹⁵ Das & Schroeder (2021); De Gregorio & Stremmlau (2021).

¹¹⁶ Berger *et al.* (2023).

¹¹⁷ Post (2009, p. 130).

¹¹⁸ Santini *et al.* (2023).

¹¹⁹ Jain (2023); MeitY India (2023); Oladapo & Ojebode (2021); Roy (2024).

¹²⁰ Oladapo & Ojebode (2021).

¹²¹ Bradshaw & DeNardis (2022).

In the case of the news media, investigation of perceptions of bias in news media coverage, beliefs in the independence of journalists, and attempts by news media to build trust concluded:

That there is *no silver bullet* for cultivating the public's trust. Audiences have different preferences around what they are looking for from news organisations and what they would personally prioritise. Our findings also reveal the degree to which attitudes about many of these strategies are dependent on existing cleavages social, political and cultural – which are also *unique to some degree to each country*. Between the Global North and Global South ... different demographic segments have distinctly different relationships with civic institutions, which shapes people's expectations towards news media as an institution as well.¹²²

Another study covering several countries found that, of those expressing skepticism about the use of algorithms in search engines and social media platforms for news selection of news, only 30% agreed that having stories selected on the basis of their previous consumption was a good way to get news, and only 27% said they favored having their news selected by news industry editors.¹²³

Research on public attitudes towards platform content moderation policies has compared attitudes to content moderation by algorithms, expert panels or juries of users, finding that expert panels were perceived to have greater legitimacy than juries or algorithms, but the findings were more significant regarding agreement with the moderation decision than they were for the process.¹²⁴

Other research in the *United States* has shown that public attitudes to the platforms' use of personal moderation tools (such as 'downvoting' posts, muting and blocking specific accounts – or, in the case of X, where individuals use Community Notes to challenge accounts whose posts they believe to be false or find offensive) versus relying on platform moderation and regulatory interventions, found, by a narrow margin, some evidence of a relationship between greater support for free speech and support for platform content moderation.¹²⁵ This might signal the public's interest in a 'pluralistic model of speech regulation ... in which speech must be regulated in a multi-stakeholder fashion' and 'approval of a shift toward a new approach to content curation that emphasizes individual choice rather than endorsing top-down censorship by platforms or other entities'.¹²⁶ It would be a mistake, however, to generalize results from the United States to other cultures and political systems where the principle of freedom of speech is tempered by the right to be free from harm.¹²⁷

Another study focused on *Brazil, India and Germany* found that interviewees agreed that mis- and disinformation 'brings to the surface and magnifies the tensions between national jurisdictions and a technology which is transnational in nature', that national regulation is 'usually toothless when problems originate beyond national borders', and that regulation may be exploited by authoritarian governments and curb competition.¹²⁸

Research highlights the need for transnational studies that go beyond the Anglocentrism present in many academic papers and reports.¹²⁹ There is also a need to reflect the views of the many regional organizations in the Global Majority World that are

¹²² Banerjee *et al.* (2023, p. 68; emphasis added), published with the support of the Meta Journalism Project. Survey N = 8,229, nationally representative on key demographic variables.

¹²³ Including Australia, Brazil, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Spain, the United Kingdom and the United States. Sample limitations meant findings are not nationally representative for India, Kenya, Nigeria and South Africa. See Newman *et al.* (2023, p. 10).

¹²⁴ Pan *et al.* (2022), supported in part by the Office of Naval Research, US.

¹²⁵ Jhaver & Zhang (2023); Procter *et al.* (2019), funded by the Economic and Social Research Council (ESRC), UK; see also Riedl *et al.* (2022) for similar comparisons. For X's Community Notes, see X (2024).

¹²⁶ Jhaver & Zhang (2023, pp. 2, 16; emphasis added), citing Balkin (2017).

¹²⁷ Rexhepi (2023).

¹²⁸ Marda & Milan (2018, pp. 10, 11); stakeholders included academia, civil society, government and policy makers, and industry including platform operators, journalists and software developers.

¹²⁹ Nguyễn *et al.* (2022), supported in part by the Siegel Family Endowment, John S. and James L. Knight Foundation and Microsoft.

commenting on mis- and disinformation governance issues.¹³⁰ It is crucial to recognize that the big tech companies have become the ‘new governors’ since they ‘are now responsible for shaping and allowing participation in our new digital and democratic culture yet have little direct accountability to their users’.¹³¹ In addition, ‘different democracies with distinctive underlying sociopolitical contexts and *cultures have varying ideas and translate values in nuanced ways regarding platform regulation*’.¹³²

As one study put it, people’s attitudes towards platform regulation are shaped by their ‘folk theories’ of partisan political agendas and commercial considerations.¹³³ A study examining ‘information disorder’ in *Asia, the Caribbean, Latin America, the Middle East, North Africa and sub-Saharan Africa* recommends a ‘critical praxis’. This would link scholarship with practical action so that ‘contextual imperatives, lived experience and local knowledge feed into policy measures’.¹³⁴ Ethnographic research in unconnected or poorly connected communities is needed to understand how information spreads through oral networks, and how mis- and disinformation manifest in everyday life.¹³⁵

Taking into account the histories of power relations and avoiding epistemologies that reduce complex understandings of ‘truth’ to a search for news media objectivity are seen as essential. So, too, is mobilizing civil society to lobby tech companies to do more to tackle information problems. This is imperative to reduce unevidenced responses to mis- and disinformation by states in the face of crises, including war and conflict.¹³⁶

Scholars argue for a multistakeholder dialogue on measures to achieve inclusive debate on responses to mis- and disinformation.¹³⁷ State-initiated mis- and disinformation strategies must be developed and implemented in an inclusive way, with an emphasis on the ‘procedural fitness’ of multistakeholder constructive and inclusive dialogue.¹³⁸ This means an obligation to consult marginalized and vulnerable populations, and to devise solutions to the challenges of mis- and disinformation based on their judgments and in ways that enable participatory action that leads to digital self-determination. Dialogue of this kind is crucial if reimagined pathways to ensure that principles, including feminist ones, inform governance strategies and practices that address the interests of vulnerable people.¹³⁹

There is evidence that governance measures could become more consistent with human rights standards if policy makers ask themselves whether legislation or non-governmental methods are the best approach to countering the harms of mis- and disinformation. When evidence is sought to underpin policy decisions, attention needs to be given to how mis- and disinformation are defined, whether there is independent oversight of interpretations of definitions and adjudication processes, and whether content governance initiatives are compatible with a healthy information ecosystem.¹⁴⁰ In general, research on governance measures aimed at moderating online content and their effectiveness identifies a lack of clarity on definitions and conflicts over values as well as divergent views about whether governance measures should be discretionary or mandatory.¹⁴¹

¹³⁰ Chairpersons Organizations (2023); UNESCO (2023c).

¹³¹ Klonick (2017, p. 1), funded by the Oscar M. Ruebhausen Fund, US.

¹³² Chung & Wihbey (2024, p. 4546, emphasis added), surveys in Mexico, South Korea, the United Kingdom and the United States.

¹³³ Mont’Alverne *et al.* (2023), N = 132, funded by the Meta Journalism Project, based on focus groups in Brazil, India, the United Kingdom and the United States.

¹³⁴ Wasserman (2022, p. 21).

¹³⁵ Wasserman (2022).

¹³⁶ Gagliardone *et al.* (2015); Haas & Kettemann (2024).

¹³⁷ Marda & Milan (2018).

¹³⁸ Marda & Milan (2018).

¹³⁹ IT for Change (2023a, b), independent NGO, Bengaluru, India. Some of these issues are discussed in Chapter 8.

¹⁴⁰ See Matasick *et al.* (2020); CNTI (2024), supported by Craig Newmark Philanthropies, the John D. and Catherine T. MacArthur Foundation, John S. and James L. Knight Foundation, Lenfest Institute for Journalism and Google.

¹⁴¹ Pradhan (2021).

6 Chapter Summary

This chapter has examined measures taken by civil society fact-checkers, big tech companies and states, including through co-regulatory approaches, to counter mis- and disinformation and to assess their effectiveness. Human rights protections are jeopardized by the way some of these measures are implemented. Many of these measures are new, and countries are at different stages in devising and implementing legislation. Evidence of their effectiveness is sparse and sometimes contradictory. This also applies to content moderation approaches, AI systems regulatory initiatives, efforts to encourage media freedom and independence, and initiatives to govern the production and use of data.

Much of the evidence that is available comes from international collaborative studies, some of which are independent. There is little clarity about the most crucial intervention points where governance can have positive impacts. Governance around data collection and the parties involved in storage, sharing and selling data, and its use in data-intensive services and products, needs to be strengthened.

The synthesis of research in this chapter shows that:

- A single approach to combating mis- and disinformation is neither feasible nor desirable. There is no 'best' policy. Current research is too focused on information itself and on technical tools to support countermeasures rather than on developing holistic investigations, acknowledging the diverse contexts in which policies and practices operate.
- Research on governance measures aimed at moderating online content and their effectiveness highlights a lack of clarity of definitions of mis- and disinformation used by researchers, especially when definitions are employed beyond the Global North. There are also conflicts over values and differing views about whether governance measures should be discretionary or mandatory.
- There is some evidence that content governance methods have a limited effect on people's responses to mis- and disinformation. For example, fact-checking is not a static process whose effectiveness can be assessed without reference to changes in techniques and practices over time.
- Many fact-checking organizations are committed to retaining a 'human-in-the-loop', but automated fact-checking may be inevitable given the rise in the volume of mis- and disinformation. Large language models (LLMs) may be a useful tool for detecting LLM-generated mis- and disinformation, but this depends on developers' motivations.
- The big tech companies could play a greater role in amplifying trustworthy information, but because these companies aim to promote their own growth, the impacts of their strategies on communities of color, women, religious minorities and LGBTQ+ people are ignored.
- Independent fact-checking, AI systems governance, news media industry regulation and data governance approaches differ regarding what is designated as harmful mis- and disinformation, who decides, and what consequences follow from those designations; and there is always a risk that enforcement leads to the suppression or violation of human rights.
- Democracies with high levels of press freedom are likely to take an holistic approach to countering online mis- and disinformation by focusing on the integrity of their election processes, news media and education. Legal safeguards are needed to prevent political influence.
- A shared understanding of the appropriate balance between the imperatives of economic growth, innovation and human rights in designing governance to combat mis- and disinformation is essential to guard against the risks of regulatory overreach by governments, particularly by authoritarian governments.

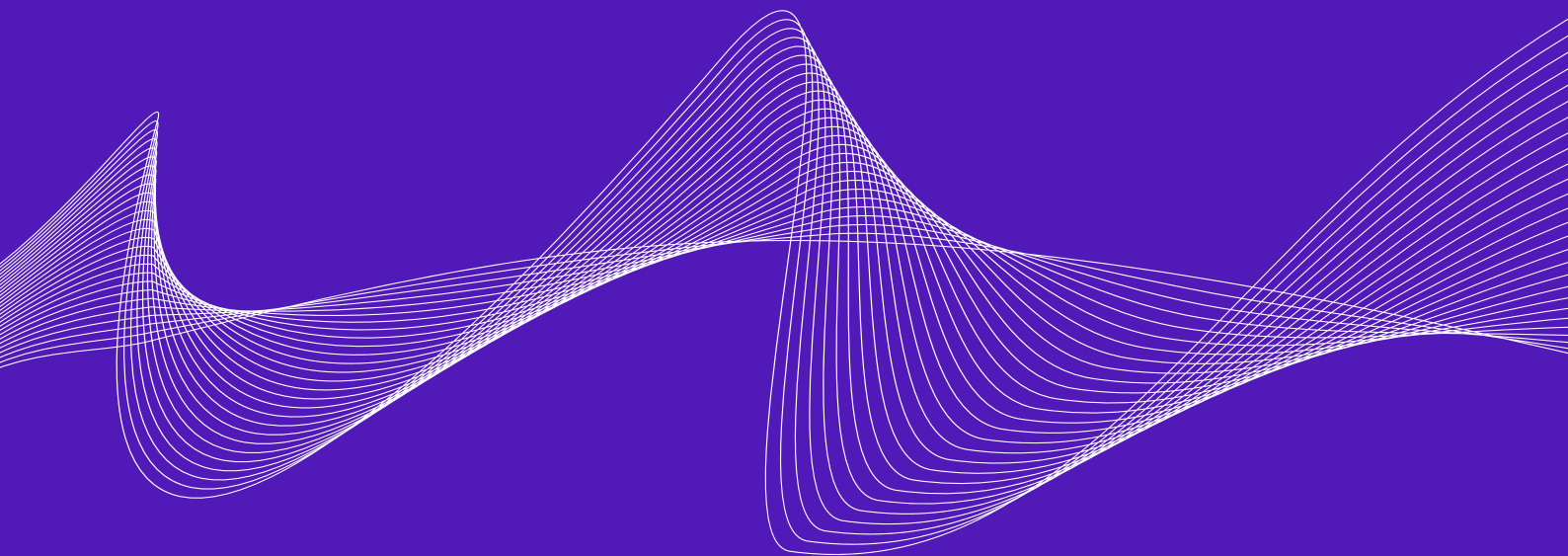
- Evidence on the public's appetite for policy interventions to counter illegal and harmful mis- and disinformation suggests a slight preference for content moderation undertaken by online participants over content moderation by platforms or state regulation, but this evidence is not conclusive or inclusive of all countries.
- To assess the extent to which civil society and other actors are underrepresented in or excluded from institutional choices about governing information ecosystems at all levels.

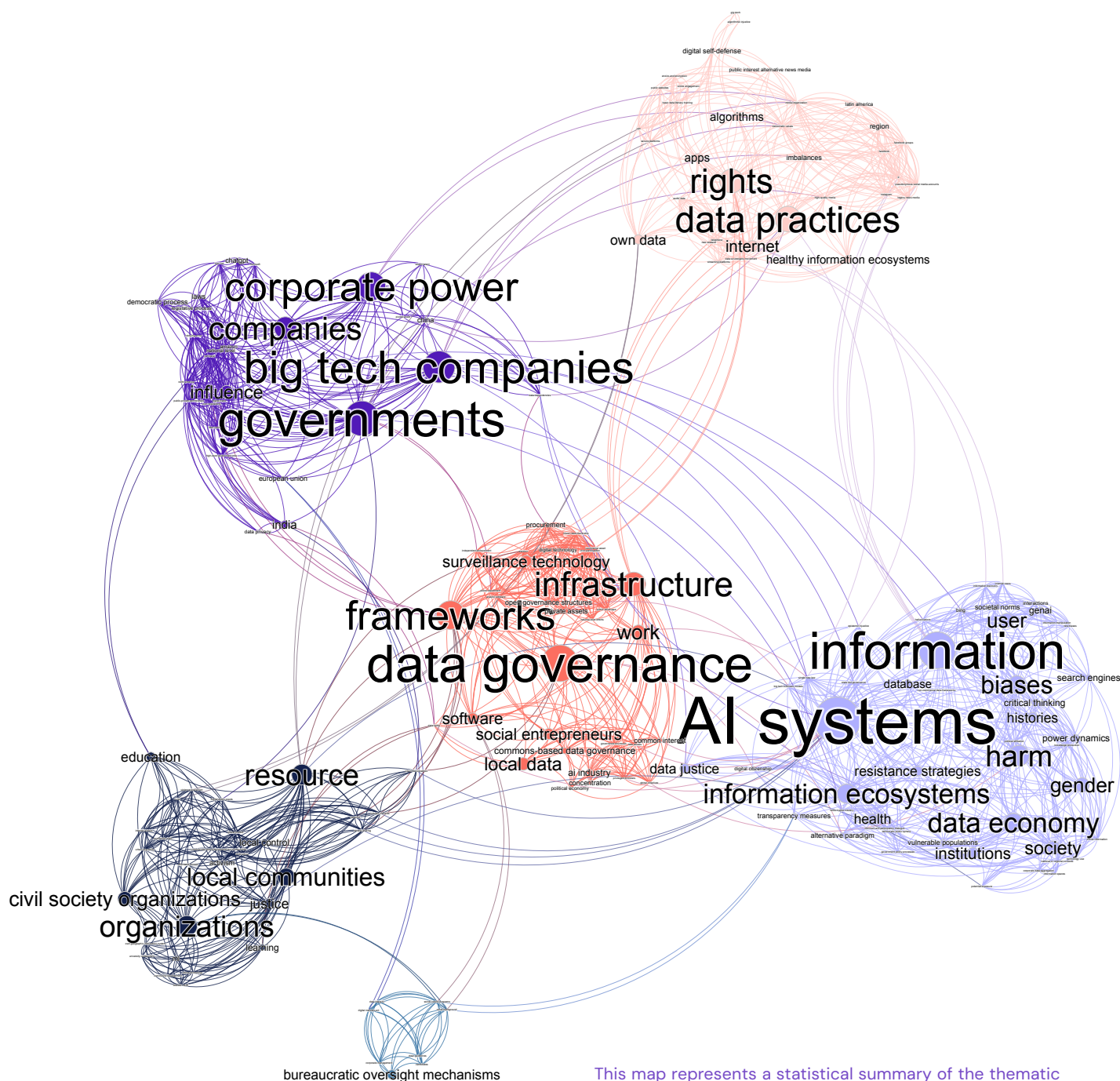
Research is needed:

- To test countermeasures to mis- and disinformation with real-world data extending beyond Global North countries. Mixed-method quantitative and qualitative approaches are essential to reveal a range of experiences. The Eurocentrism present in research needs to be addressed to make visible the views of the many organizations in the Global Majority World that are working on mis- and disinformation governance issues.
- To examine the effectiveness of existing rules applying to automated content governance systems using risk assessments and audits.
- To examine whether content governance measures are too general to be complied with, to identify shared terminology and to create frameworks for company provision of the granular platform data that is needed for independent assessment of these measures.
- To investigate what content governance practices would be acceptable to online participants in different countries, and what countermeasures would be effective in reversing declines in public trust in public institutions and the news media in countries where this is occurring.
- To monitor digital platforms' practices including abandoning content self-regulatory measures, laying off staff, weakening privacy policies and imposing limits on fact-checking.
- To examine the incidence of content and other governance measures that result in the suppression of voices that are critical of state authorities.

CHAPTER 8

TOWARDS DATA JUSTICE IN INFORMATION ECOSYSTEMS





This map represents a statistical summary of the thematic content of this chapter. The network graph represents relations between the words in the chapter, placing them closer to each other the more they are related. The bigger the node, the more present the word is, signalling its role in defining what the report is about. The colors represent words that are closely related to each other and can be interpreted as a topic.

The map is generated by the OID on the basis of the chapter's text using GarganText – developed by the CNRS Institute of Complex Systems. Starting from a co-occurrence matrix generated from chapter's text, GarganText forms a network where words are connected if they are likely to occur together. Clustering is conducted based on the Louvain community detection method, and the visualization is generated using the Force Atlas 2 algorithm.

[Link to the interactive map here](#)

This chapter examines how the monopolistic power of big tech companies – permitted by state and co-regulatory measures and pursued by big tech companies – creates biases and harmful discrimination and exclusions, infringes on people's human rights in a data economy that thrives on data extraction and monetization, and diminishes the health of information ecosystems.¹

The research synthesis in this chapter focuses on:

- **Why do corporate incentives, strategies and practices involved in designing, developing, selling and controlling data lead to epistemic injustice?** Recent histories of digital innovations and their impacts in the Global North and in the Global Majority World are critically examined. We explain how corporate practices create dependencies and restrict people's abilities to control how datafication impacts their lives, as well as the need for individual and community control, autonomy and authority if the struggle to achieve healthy information ecosystems is to succeed.
- **What strategies and tactics are individuals and communities developing to resist the extractive features of the data economy?** We discuss how individuals and groups are working to reimagine and implement data governance frameworks, practices and technical designs that could yield healthier information ecosystems, combat mis- and disinformation and improve prospects for democratic participation.

This chapter emphasizes the individual and collective dependencies and inequities that result from datafication, and how datafication practices can be reimaged to empower individuals and communities in the data economy and contribute to data justice.

Chapter 9 summarizes insights from the synthesis of research in this report. For key highlights, see the Executive Summary.

¹ For background reading, see Casati (2013); Couldry & Mejjas (2019); Fuchs (2023); Hintz *et al.* (2019); Mejjas & Couldry (2024); Papacharissi (2013); Powell (2021); Söderström & Datta (2023); Vaidhyanathan (2011, 2022); van Dijck *et al.* (2018a). See Appendix: Methodology for details of literature review process.

1 Introduction

We begin with a discussion of the phenomenon of epistemic privileging. How do knowledge, structures and practices combine with technological affordances to shape what counts as good knowledge and practice? Whose knowledge matters when it comes to governing information ecosystems?² Current information ecosystems and their data governance arrangements impact on the quality of information, on the risks and realities of reproducing or worsening socio-economic, gender, racial and other forms of discrimination, partly because of the biases in AI models. These developments lead to distortions in understanding and in decision-making, diminishing the health of information ecosystems, and especially the conduct of debate in the public sphere.

Epistemic privileging that is distinguished by class (caste), race, gender, political affiliation and economic status is not new.³ The biases and exclusions in the histories of technological innovations and their impacts in the Global North and in the Global Majority World always need to be critiqued, and policies, regulations and practices revised when they marginalize populations.⁴ This chapter examines the limitation of solutions designed to tackle problems such as discrimination and bias in the performance of AI systems. It does so by recognizing that information – however produced – is always interpreted in the light of power structures. Epistemic privileging of someone's knowledge is inevitable because information and knowledge are not neutral.⁵

The inequitable outcomes of epistemic privileging cannot be addressed merely by balancing priorities

for investment in digital technology or by prohibiting specific applications, for example facial recognition technologies. Instead, a profoundly deeper understanding is needed of how historical and contemporary power dynamics shape technological development and deployment, reinforcing entrenched inequities that influence which voices are heard and which ones are silenced. Knowledge is inextricably linked to power, and the *control* over knowledge production and dissemination can reinforce existing hierarchies, as is acknowledged in critical research on how societies come to be governed.⁶

Critical examination of these power dynamics reveals that no configuration of technological affordances can be universally beneficial, and there are many ways these affordances can perpetuate disparities.⁷ Global efforts to increase reliance on data drive perverse economic incentives when marketeers operate to capture people's attention in a data economy that thrives on data extraction and monetization.⁸ It is essential to move away from perspectives emphasizing 'data universalism' and assuming a homogeneous experience of datafication across the world. A critical perspective on the extraction, accumulation and commodification of data and how this influences people's lives is a necessary step in understanding and resisting unjust power dynamics.⁹ The history of 'data capitalism' is of a socio-technical system that results in a 'distribution of power that is asymmetrical and weighted toward actors who have access and the capability to make sense of data'.¹⁰

Digital divides – a term describing gaps between those who have access to and can benefit from modern digital technologies and those who do not and cannot – illustrates the material consequences

² Wihbey (2024); Wu (2017).

³ Horowitz *et al.* (2024).

⁴ Chambers (1997); Thakur & Madrigal (2022); Willems (2014b).

⁵ Rouvroy & Berns (2013).

⁶ Foucault (1980).

⁷ This point is made consistently in relation to the consequences of illiberal regimes (Sodré, 2021), and in relation to the need for data activism to support counter-epistemic and alternative practices (Segura & Waisbord, 2019).

⁸ Misra (2022).

⁹ Milan & Treré (2019), supported by the European Research Council (ERC) and Horizon program; Cieslik & Margócsy (2022); Horst *et al.* (2024), supported in part by the Australian Research Council (ARC); Arriagada *et al.* (2023), supported in part by IDRC (International Development Research Center), Canada and by the Millennium Nucleus on the Evolution of Work. For a study of how people's everyday lives are affected, see Dunn *et al.* (2024), with case studies of resistance to 'algorithmic authority' in Argentina, Brazil, the Caribbean, China, Ghana, India, Jamaica, the Philippines, Russia, South Africa and Southern Africa and the United Kingdom (Domingos Cordeiro & Cozman, 2024).

¹⁰ West (2019, p. 23).

of epistemic privileging in terms of outcomes for individuals and groups. Overcoming access divides and differences in capacities to interpret online information alone is insufficient, however, for achieving digital equity. What is required is attention to the social, cultural and political institutions that either hinder or facilitate meaningful and beneficial uses of technology. This necessitates policies and practices that address the socio-economic and cultural barriers that hinder equitable access and technology use.¹¹

Algorithmic bias has been a feature of computational systems for decades, but its more recent manifestation in today's information ecosystems creates newer forms of epistemic privileging as new divides enabled by digital platforms, social media and search engines and a host of other AI applications reinforce racial and gender stereotypes, privileging certain perspectives over others.¹² This is not an incidental byproduct of the progressive innovation in and adoption of digital technologies, but a reflection of the socio-political contexts within which these technologies are developed and deployed.

Earlier chapters noted the dominance of scholarship on unhealthy information ecosystems, which is contextualized by the experience of those in the Global North.¹³ In this chapter, we take a step towards decolonizing knowledge about information ecosystems so that the experience of those in the Global Majority World might inform choices about how information ecosystems should be governed. The aim is to reimagine information spaces that privilege fairness, justice and human rights, not just in principle, but in practice.¹⁴

2 Strengthening Deliberation and Democracy

The burdens of corporate data aggregation in today's data economy are disproportionately borne by those who are, or historically have been, subject to forms of social, economic, political and cultural inequality or oppression, and government policy making tends to exclude these same groups.¹⁵ Even in local, national or regional contexts, where there are laws, policies and practices designed to promote participatory politics and democratic self-governance, explicit policy-making processes are typically top-down and controlled by powerful elites, even when they are structured to perform as 'representative' government.¹⁶ Those who are most vulnerable to the potential exposure of their information and to injustices and inequalities that come with massive data aggregation are least well-positioned to seek and obtain remedies for individual harms or to participate in civil society advocacy on these issues.¹⁷ The space available for political communities to push for legislative reforms that might limit or shift data practices in fundamental ways is diminished as industries and bureaucracies become more dependent on digital infrastructures and algorithmic products.¹⁸

2.1 CORPORATE POWER AND INTERESTS

Political processes involving government institutions relating to data governance are heavily influenced by corporate interests. This influence takes the form of lobbying, whereby powerful corporations engage highly paid professional lobbyists to meet with government officials and others who are well

¹¹ This report does not discuss the complexities of digital divides directly, but see Warschauer's (2004) early work in the United States and more recent studies on digital inclusion and outcomes, for example, Helsper (2021); Robinson *et al.* (2020), supported in part by FONDECYT (Fondo Nacional de Desarrollo Científico y Tecnológico [National Fund for Scientific and Technological Development]), Chile, the National Agency for Research and Innovation (ANII, Agencia Nacional de Investigación e Innovación), Uruguay, Social Science and Humanities Research Council (SSHRC) of Canada, and the Internet Society; see also Gillwald & Weleilakeba (2024); Heeks (2022); Ragnedda & Ruij (2020); Schaafe & Fukuyama (2023); Trappel (2019); Yates & Carmi (2024); Hargittai (2021).

¹² Noble (2018).

¹³ Schoon *et al.* (2020).

¹⁴ Alaimo & Kallinikos (2024); Gillwald *et al.* (2022); Gurumurthy & Chami (2024); Mejias & Couldry (2024); Santos & Ndlovu (2022).

¹⁵ Glasberg & Shannon (2010); Glimmerveen *et al.* (2022); OHCHR (2014).

¹⁶ Glimmerveen *et al.* (2022); Wike *et al.* (2024).

¹⁷ Broomfield & Reutter (2022); Eubanks (2018); Georgiou (2023); O'Neil (2016); Ross Arguedas & Simon (2023); Trappel (2019).

¹⁸ Mager & Katzenbach (2021); Papaevangelou (2023); Whittaker (2021).

positioned to shape law-making.¹⁹ An example is the legislative process that resulted in the passage of the European Union's AI Act. In the lead-up to the March 2024 adoption of the Act, European AI startups, such as Aleph Alpha and Mistral, as well as American tech giants, including Google, Microsoft and OpenAI, lobbied aggressively for amendments to the draft legislation that would favor their own products and corporate practices.²⁰ European watchdog organizations reported that tech companies had 'privileged and disproportionate access to high-level European decision-makers'.²¹ OpenAI, in particular, lobbied European Commission officials to ensure that 'general purpose' algorithmic models, such as the one underlying OpenAI's chatbot, ChatGPT, would not be treated as 'high risk' by default under the new legislation.²² In the United States, the largest tech companies spent close to USD 70 million lobbying in both 2022 and 2023. Much of this went towards influencing United States federal policy, but tech companies also make considerable investments in lobbying to limit the scope of state-level privacy legislation, which, in the context of many years of congressional gridlock, has been active political terrain when it comes to data privacy.

Investigative reporters and civil society organizations uncovered what they characterized as 'a coordinated, nationwide campaign by Big Tech' to shape state-level privacy laws.²³ Of 14 state-level privacy laws, 'all but California's closely follow a model that was initially drafted by industry giants such as Amazon'.²⁴ Companies also have a long track record of exploiting public emergencies to engage in turbo-charged lobbying that evades the democratic process.²⁵ During the Covid-19 pandemic, Eric Schmidt, former CEO of Google, leveraged the panic of citizens and policy makers

to campaign for massive public expenditure on Research & Development, and for the creation of dozens of 'public-private partnerships' to embed data-intensive, corporate-owned tech platforms across multiple sectors – notably healthcare and education.²⁶

The lobbying arms of global tech firms also spend significantly to influence policy at the national level in countries outside Europe and North America, especially where emerging data markets promise to be very large. In *India*, for example, Meta, Google and Amazon lobbied aggressively against data localization provisions in the country's data protection bill.²⁷ When *Brazil* introduced legislation to combat 'fake news' in 2023, American-based tech companies campaigned against the bill. Google used its search engine home page to promote articles criticizing the legislation and urging Brazilians to act against it.²⁸

Tech companies also seek to influence legal frameworks that do not specifically target digital products and services, but which impact on profit margins by lobbying policy makers on issues related to trade and the global economy. For example, during negotiations around the Indo-Pacific Economic Framework, big tech companies pushed for specialized 'digital trade' rules that would have limited the ability of 14 countries to enact their own regulations intended to restrain the activities of tech companies at the national level.²⁹

Tech companies wield influence over policy-making processes through entrenched relationships with policy makers in another way. This takes the form of a 'revolving door' between government and industry, whereby people leave high-level government jobs for high-paying corporate jobs, and vice versa.³⁰

¹⁹ Bannerman *et al.* (2024); Popiel (2018); Rankin (2023); Ruohonen (2003); Tarrant & Cowen (2022), supported by the Social Sciences and Humanities Research Council (SSHRC) of Canada.

²⁰ Corporate Europe Observatory (2024b); Perrigo (2023).

²¹ Corporate Europe Observatory (2024a).

²² Perrigo (2023).

²³ Feathers & Ng (2022).

²⁴ Feathers & Ng (2022); Fitzgerald *et al.* (2024).

²⁵ Klein (2008).

²⁶ Klein (2020).

²⁷ Business Standard (2018); Sherman (2022).

²⁸ Boadle (2023); Harris (2023).

²⁹ Birnbaum & Martin (2023); James (2022); Lawder (2023).

³⁰ Alfonsi (2019); Popiel (2018).

These longstanding relationships become the context in which informal lobbying through socializing takes place, which can effect legislative and regulatory outcomes.³¹ The porous boundary between government and industry also results in more pervasive and difficult-to-document forms of influence, such as ‘corporate influence on regulators’ systems of belief, policy preferences and ideological biases’.³² This kind of influence can take years to manifest in concrete policy and is rarely discernible in any single legislative process.³³

Another kind of industry–government relationship that tightens the corporate grip on policy making is when companies have a monopoly on a set of services that government requires. One example is the growing dependency of national security infrastructures on surveillance technology developed and maintained by major tech companies such as Palantir.³⁴ As whole sectors within state social, political and economic systems become dependent on the data-intensive products and infrastructures developed and sold by corporations (for which they claim both the rights of intellectual property and the protection of trade secrets), the kinds of limits that policy makers are willing to impose on corporate data practices become narrower.³⁵ These developments are politically charged and extend from services to infrastructure, including undersea cables that have enabled the *United States* to surveil other countries and are now facilitating countries like *China* and *Russia* to do the same, as struggles over the ‘underground empires’ ramp up.³⁶

While much attention focuses on dependencies and inequalities associated with American or Chinese-owned big tech companies, regional companies also create dependencies and operate with extractive data economy models. For example, Mercado Libre

is an Argentinian platform company (the fourth largest Latin American company) that engages in electronic commerce, fintech payments and credit assessment. It is dependent on the cloud services of Amazon and Google, but it also develops inhouse data services that exploit users in the Latin American region.³⁷ This highlights the multilayered and global nature of data dependencies and their complexities.

Data dependencies are also created in other ways that impact on how data is collected, stored, shared and used. An example is the procurement process through which government agencies acquire or license data-intensive products developed by companies that are then incorporated into public systems.³⁸ These processes can be opaque, and even when transparent, they do not usually include a mechanism for public comment or participation. Over the last 30 years, vast digital bureaucracies have been put in place through procurement processes without the knowledge of those whose lives are most affected by them, and often without even the knowledge of elected officials who are supposed to represent their constituents’ interests. Procurement processes leading to the creation of digital bureaucracy profoundly change the way people relate to their government, in some cases imposing substantial hurdles on those trying to access services and benefits to which they are entitled.³⁹

Treating data as an economic good, even a public good, that generates revenue and profits is hardwired into today’s information ecosystems. It results in endless amounts of data monetization as data becomes a pivotal asset in the data economy, but it is widely criticized for failing to deliver other societal goals.⁴¹ The AI industry that depends on data extractivism (the large-scale harvesting of data by private companies) to build its systems is

³¹ Li (2023).

³² Popiel (2018, p. 568) and see Caplan (2023); Pickard (2014); Teachout & Khan (2014).

³³ Popiel (2018).

³⁴ Ball & Snider (2013); Iliadis & Acker (2022, pp. 334–363); Popiel (2018); and as noted in Chapter 7.

³⁵ Singh & Gurumurthy (2021).

³⁶ Farrell & Newman (2023).

³⁷ Franco *et al.* (2024), funded in part by the National Scientific and Technical Research Council (CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas), Argentina.

³⁸ Calo & Citron (2021); Crump (2016); Faife (2022).

³⁹ Crump (2016); Faife (2022); Hardy & Williams (2008).

⁴⁰ Luchs (2023).

⁴¹ Purtova & van Maanen (2024).

dominated by a few powerful companies (mainly in China and the United States). This concentration of power is supported by disproportionate access to resources, including computing power, high-quality data and expert talent.⁴² This dominance disadvantages smaller entities and independent innovators. The vast resource requirements for competitive AI systems development — such as advanced computing infrastructure and large-scale data sets — are often beyond the reach of smaller developers and firms, perpetuating the dominance of large tech companies. Talented AI professionals are drawn to large firms offering better financial and career incentives, resulting in talent consolidation and excluding smaller players and many academic institutions.⁴³ The high cost of entry and not having access to proprietary data sets disadvantages those seeking to develop supportive community data governance frameworks and practices.

2.2 DATA, AI SYSTEMS AND DISCRIMINATORY BIAS

Rapid advances in AI systems and their integration into search engines and conversational applications such as ChatGPT yield benefits and potential harms in terms of the quality of information.⁴⁴ Described as ‘revolutionary’, these technologies embody inherent risks related to algorithmic bias and information manipulation that impact on people’s decisions and on societal norms.⁴⁵

AI models, including those powering generative AI (GenAI), such as ChatGPT, incorporate biases from their training data, and can develop new biases through interactions with their users.⁴⁶ These biases manifest in several harmful ways. They can surface in the reproduction of existing biases and through their impact on user perceptions. Just as biases are present in human-generated content, large language models (LLMs) can inadvertently

perpetuate these biases, leading to a skewed representation of facts and socially constructed biases in AI-generated content. These biases then can shape how individuals perceive reality, potentially reinforcing stereotypes or presenting a biased view of events and histories; they are an opaque form of epistemic privileging.⁴⁷

2.2.1 Reproducing Bias in AI Models

AI models trained on data, including GenAI systems, reflect the biases present in data they are trained on, or biases developed through users’ interactions with them. These can skew user behavior, potentially perpetuating stereotypes and generating mis- and disinformation. If an AI system is trained on historically biased data, it may generate responses that are subtly prejudiced, reinforcing harmful norms instead of challenging them.

Information discovery for internet users is conducted using search engines such as Google, Yahoo or Bing. Until recently these presented their results to users as a list of sources (ranked according to search engine-defined ‘relevance’ criteria, influenced by advertising expenditure). They have long since moved beyond simply counting the number of in-bound links to a webpage. Users can still choose which links to follow, but studies find a search engine manipulation effect (SEME). By altering search result rankings or manipulating result visibility, this may influence consumer choices, and even voting behavior.⁴⁸ Google and Bing use GenAI to create conversational search assistants that summarize search results instead of simply listing them.⁴⁹ This may reduce a user’s ability to discover diverse viewpoints, potentially limiting exposure to multiple viewpoints, diminishing abilities for critical thinking and source evaluation skills, and decreasing agency over the information that is consumed.⁵⁰

⁴² Luchs (2023).

⁴³ MIT Technology Review Insights (2023).

⁴⁴ See Chapter 3 for a discussion of bias.

⁴⁵ Ferrara (2024b).

⁴⁶ Ferrara (2024a).

⁴⁷ Machill (2020).

⁴⁸ Epstein & Robertson (2015), supported by the American Institute for Behavioral Research and Technology (AIBRT), a non-profit organization, US.

⁴⁹ Microsoft has launched Copilot, which integrates ChatGPT into its Bing search engine. Google had added Gemini to its search tool at the time of writing.

⁵⁰ Hadi Mogavi *et al.* (2024).

GenAI systems may also unintentionally misinform users due to what are known as ‘hallucinations’ (plausible responses that have no basis in reality), and that scholars argue are either a natural consequence of the underlying technology or of errors due to bias in the training data.⁵¹ This can lead to confusion and the spread of mis- or disinformation. When GenAI systems are used to target certain users with specific content, this has the potential to subtly influence opinions and behaviors, potentially unethically.⁵² This poses a risk of deliberate or voluntary manipulation, and is said to require the introduction of guardrails.⁵³

While AI technologies offer unprecedented access to information and the ability to analyze vast data sets, they also require careful management to mitigate risks associated with bias and manipulation. Moving forward, developers and policy makers must collaborate to implement robust ethical guidelines and transparency measures to ensure that advances in AI contribute positively to society without compromising the integrity of information. As emphasized in Chapter 3, all biases cannot be eliminated, and the potential for unfavorable treatment of individuals or groups remains.

2.2.2 Inaccuracies and Distortions in Decision-Making

Bias is an element of human cognition that can serve as a heuristic for faster decision-making in complex environments.⁵⁴ When embedded within AI systems, however, biases can perpetuate harm, especially when they inflict unfavorable outcomes on individuals or groups. The consequences of decision-making based on flawed or biased data are potentially substantial, and include financial loss, reputational damage and legal penalties. Bias

can manifest as direct discrimination where AI systems provide less favorable results to users from certain demographic groups.⁵⁵ When these models are used as the decision-maker for social justice programs, for example, this can lead to the exclusion of marginalized groups, and they have been shown to worsen existing inequalities and diminish trust in data-driven systems.⁵⁶ Similarly, hiring algorithms can prefer candidates of a specific gender or racial background irrespective of their qualifications.⁵⁷

To mitigate these risks accompanying datafication using AI, organizations need to implement comprehensive data governance frameworks with clear guidelines for data quality, usage and security.⁵⁸ This includes developing standardized procedures for data collection, validation and storage, and using software tools that manage data quality and can be used to correct inaccuracies in real time.⁵⁹ Regular bias audits and algorithm reviews are crucial. These can be facilitated by third-party auditors and the use of fairness tools in machine learning (ML) to adjust models.⁶⁰ Reforms to data governance frameworks and practices can help to improve data integration and address bias by adopting modern data architecture principles, and implementing enterprise data management platforms that can make it easier to handle large data sets. By understanding the sources and impacts of data analytics flaws and employing effective mitigation strategies, organizations can improve data integrity and decision-making accuracy. This is essential for ensuring that AI systems are used responsibly and ethically, promoting equity in automated environments.⁶¹

Bureaucratic oversight mechanisms like auditing and litigation have had a limited impact on the power wielded by big tech companies. In the European Union, the regulatory package (Digital

⁵¹ Xu *et al.* (2024).

⁵² Motlagh *et al.* (2023).

⁵³ Hao *et al.* (2024); Linehan *et al.* (2024), authors members of Object Management Group, an industry standards consortium, IBM and the Industry (IoT) Internet of Things Consortium.

⁵⁴ Kahneman *et al.* (2021).

⁵⁵ Angwin *et al.* (2016).

⁵⁶ Park & Humphry (2019).

⁵⁷ Raghavan *et al.* (2020).

⁵⁸ For a review of data governance frameworks, see Marcucci *et al.* (2023), supported by the World Health Organization.

⁵⁹ Veiga *et al.* (2017).

⁶⁰ Verhulst (2024).

⁶¹ Mitchell *et al.* (2021).

3 Alternative Data Governance Practices

Services Act and Digital Markets Act) has created new enforcement mechanisms that can be used to compel corporate transparency concerning corporate management of data, and to impose major penalties on companies that fail to comply with the legislation.⁶² It is unclear whether actors will have the resources and political will to use these mechanisms in a way that substantially changes the big tech companies' monopolistic practices when it comes to their treatment of data in Europe or beyond. The European Commission's task force with responsibility to 'check whether some of the world's richest tech firms are complying with rules designed to make them cede some ground to their smaller competitors' is under-resourced compared to the big tech companies.⁶³ In the United States, while there is no federal legal framework regulating corporate data practices, there are legal frameworks relating to fair labor, anti-trust, intellectual property and contract law that could be used to challenge the dominance of the large tech companies. Until recently there has been little sign of significant enforcement activity against tech companies, and when cases are brought and even won, they involve very lengthy proceedings.⁶⁴

Since the use of data is pervasive in the modern data economy, attention is being given to alternative data governance norms, architectures, institutions and practices to enable individuals and groups to gain greater control over data, potentially increasing information integrity and the health of information ecosystems.

Political engagement that takes place in parallel to, in resistance against, or separately from, government policy processes and existing data governance legislation is especially important if data justice is to be achieved.⁶⁵ There are many examples globally of communities organizing to resist datafication, to develop alternative local data practices (including alternative governance principles specific to local practices), and to establish systems for creating and sharing knowledge and information (digital or otherwise) that do not depend on the digital products owned by large companies.⁶⁶ These do not necessarily represent (or aspire to be) scalable forms of resistance to big tech monopolization of digital infrastructures or dominance in information ecosystems. However, they offer models for how communities can engage in democratic contestation, responding to questions that arise about data and digital infrastructure in established forums offered by governments, and raising questions about whether such forums are asking the right questions. This section explores what 'digital democracy' and data justice might look like.⁶⁷

Resistance strategies and practices try to embrace an obligation to consult marginalized and vulnerable populations, to devise solutions based on their judgments and to enable participatory action that lead to digital self-determination.⁶⁸ Figure 8.1 locates people and their communities aiming to strengthen data justice and to create the potential for inclusive, informed and participatory dialogue in a democratic public sphere at the center of

⁶² For details of European Union regulatory package, see Chapters 6 and 7.

⁶³ Hancock (2024).

⁶⁴ Landau (2021).

⁶⁵ Data justice is addressed in work by Milan *et al.* (2021), funded by the Nuffield Foundation; Niklas & Dencik (2024), supported by the European Research Council (ERC).

⁶⁶ Examples are given by Bhat (2021); Carroll *et al.* (2019); Dutta & Pal (2020); Mejias & Couldry (2024); The Tierra Común Network (2023). A series of reports produced by the Institute of Development Studies (IDS), Sussex, provides profiles of the digital rights landscape in African countries (Cameroon, Egypt, Ethiopia, Kenya, Nigeria, South Africa, Sudan, Uganda, Zambia and Zimbabwe), as of 2021; see Roberts & Ali (2021).

⁶⁷ Ford (2019). Approaches to 'deepening democracy' through participatory governance have a long history of discussion in the literature; see Fung & Wright (2003).

⁶⁸ Medrado & Verdegem (2024); Zhang *et al.* (2023).

information ecosystems. This can be achieved by contesting corporate power, by opposing the biases created by the outputs of AI systems and moving towards collective or public ownership of data, depending on what alternative is most appropriate in the given country. Alternatives are available for the news media industry, for the development of AI systems and for governing data in the common interest rather than in the interests of big tech companies. The rest of this section discusses a variety of resistance strategies, ranging from individual measures for people to defend themselves from exploitative data practices to community (and sometimes national) strategies to change the way data is governed.

The challenge is to address a ‘double helix’ of extraction, whereby data is extracted from places and people in the Global Majority World by the technology industry, and knowledge about this process is extracted by researchers and their institutions in the Global North. Researchers in the Global North (or in positions of power in the Global Majority World) need to be sensitized to the voices of their research participants if alternative approaches to data governance are to be imagined and put into practice to represent the needs of marginalized people, instead of reproducing epistemic injustices through patterns of data and knowledge extraction.⁶⁹

One aspect of such resistance strategies involves defining what empowering digital citizenship might involve. In the Global North, there is a robust literature on active citizenship and participation in society, demonstrating that it does not materialize in a vacuum.⁷⁰ In the Global Majority World, resistance to data extraction can mean unpacking what digital citizenship means if it is not skewed by the decisions of distant big tech companies or autocratic states.⁷¹ However, ‘one cannot simply brush away these new forms of dispossession and inequality with a single new law, a revolutionary technology or even a social revolution’.⁷² Prospects for empowering data governance can be improved by confronting data extractivism.⁷³ This can occur through unionized worker resistance, or other means.⁷⁴ Other strategies include those developed by Indigenous communities to draw attention to how the predictive power of algorithms, such as Google Search’s Autocomplete, treats gender and political keywords in languages such as Amharic, Kiswahili and Somali in ways that amplify power imbalances.⁷⁵ In the case of news media, resistance practices include building skills and developing ethical data practices or efforts to counter online ‘mob censorship’ when it threatens to silence journalists and puts their lives at risk.⁷⁶

Figure 8.1
Information ecosystems – alternative data governance approaches and resistance strategies



Source: Authors of this report.

⁶⁹ Enghel & Noske-Turner (2018); Lehuédé (2022).

⁷⁰ Hintz *et al.* (2019); Isin & Ruppert (2020).

⁷¹ Roberts & Bosch (2023b).

⁷² Mejias & Couldry (2024, p. 206).

⁷³ Graham & Dittus (2022); Graham & Ferrari (2022).

⁷⁴ Graham & Dittus (2022); Graham & Ferrari (2022).

⁷⁵ Chonka *et al.* (2023).

⁷⁶ Nechushtai (2023); Waisbord (2023).

Innovative, non-commercial data frameworks and practices can profoundly transform how data is collected, processed, stored and used to meet local community and individual needs. They have the potential to alter information ecosystem landscapes, data-related operations, societal norms and the

integrity of information within these systems.

Table 8.1 summarizes a set of corporate datafication resistance strategies and tactics for the purpose of resisting data extractivism and mobilizing new ways of practicing data governance.

Table 8.1
Corporate datafication resistance strategies and tactics

Actor	Tactic	Purpose
Individual self-defense strategies	Adopt privacy-enhancing technologies such as virtual private networks (VPNs) and encrypted messaging applications.	Digital self-defense practices and digital dissent.
	Opt out of dominant social media platforms.	
	Remove personal information from public and private data sets.	
Public interest alternative news media	Investigate corporate data practices and concentrations of corporate power, with a focus on the impacts of practices on marginalized groups.	Enhance public awareness of the harms of poorly constrained commercial datafication.
		Hold technology companies accountable to limit expansion of corporate power.
Community collaborative strategies – Indigenous communities and municipal initiatives	Produce or collect data relevant to the needs of communities.	Establish a citizen-first, rather than technology-first, approach to data governance.
	Develop community-owned and run platforms for recording and sharing information.	
	Establish community principles for data with or about communities and municipalities.	Create alternative data norms. Model the creation of new data norms for other communities.
	Demand that policy around data responds to the needs of municipalities.	Democratize legislative processes around datafication, and fight for new ordinances.
Social entrepreneurs and community-controlled technologies and data practices	Partner with local communities to develop new technologies based on non-extractive data practices.	Provide alternatives for individuals and communities to avoid contracting with multinational companies.
	Develop open-source software applications.	Provide personalized tech support and digital literacy training.
	Develop public data sets with local data capturing local knowledge.	
New national-level decentralized data governance frameworks	Build networks and own and control data.	
	Develop new data governance frameworks that preserve local control of data.	Encourage community data hubs, decentralized data infrastructures, local data analytics, data lockers and cooperatives and public data infrastructures.
	Develop commons-based data governance.	

Actor	Tactic	Purpose
Civil society organizations, researchers and philanthropic organizations	Establish systems of social support for communities harmed by tech dominance and datafication.	Mitigate harms to individuals and communities, and remove barriers to organizing and other forms of democratic participation for those most negatively impacted by datafication.
	Engage with local, national and international human rights bodies to document impacts of massive data aggregation and algorithmic outputs, aiming for social and economic justice.	Develop global and intersectional analyses of how datafication and corporate monopolization impact justice and democracy and articulate the common good.
	Develop, facilitate and support participatory action research in collaboration with and within local communities.	Build tech literacy in local communities and their capacity to use data on their own behalf, or to resist its use against their interests.
	Undertake research to expose harms of data-intensive tools and infrastructure, and identify strategies for democratic data governance.	Add to knowledge about the political economy of data. Enhance public awareness of harms of poorly constrained datafication.
	Fund individuals, organizations and institutions engaged in the activities described in this table.	Redistribute power over data, and power in decision-making about data, away from big tech companies towards the political community.
	Engage in litigation to enforce existing data governance frameworks.	Engage in litigation to enforce novel data governance frameworks.

Source: Authors of this report

Ambitions for building healthy information ecosystems depend on the agency of individuals and groups to resist the power of technology companies. Studies of this ‘contested battleground’ often undertake ethnographic research to examine how people develop strategies and tactics to resist the way algorithms influence their lives. This work demonstrates that algorithms can be appropriated by users, with examples from gig work, the cultural industries and politics. It reveals how people invent practices that – even if temporarily – enable them to transgress algorithmic systems.⁷⁷ This research tradition supports experimentation and efforts to imagine alternatives to ‘algorithmic injustice’.⁷⁸ It avoids a ‘cybernetic ideology’ that couples technology innovation with modernity and progress, assuming there is only one direction of change.⁷⁹ Research of this kind is needed to reveal novel ways of defending people’s human rights, including their epistemic rights.⁸⁰

3.1 INDIVIDUAL DIGITAL SELF-DEFENSE STRATEGIES

A basic practice of resistance available to individuals is digital self-defense, which is increasingly taught globally and practiced by activists as part of the work of political organizing. Community-based organizations are spreading digital self-defense resources to give everyone the means to protect themselves from some of the most acute consequences of data surveillance. Digital self-defense practices may include improving the security of passwords, accessing the internet through a VPN, using messaging apps employing end-to-end encryption, and removing one’s own data from public websites and – where possible – from the custody of data brokers.

Some digital self-defense curricula encourage opting out of most social media, or they suggest more secure platforms for online engagement. Basic data literacy training is often a component of digital self-defense training. While digital self-defense focuses on the personal security of individuals,

⁷⁷ Bonini & Treré (2024, p. 3).

⁷⁸ Buolamwini (2023); Cammaerts & Mansell (2020); Mansell (2012); Mejias & Couldry (2024); Noble (2018).

⁷⁹ Caballero & Monje (2024).

⁸⁰ Horowitz *et al.* (2024).

and not on systemic change-making, creating the possibility for digital security is an important precondition for larger-scale political engagement.⁸¹

3.2 PUBLIC INTEREST ALTERNATIVE NEWS MEDIA

Alternative ways of providing news have the potential to operate as a resistance strategy that improves the health of information ecosystems, although this is not always the outcome.⁸² Public interest alternative news media operates as a counter-public sphere to mainstream news media. When these media organizations are informed by respect for human rights and democratic values, they are better positioned to investigate corporate data practices, to support communities and to engage with local, national and international human rights bodies that aim to expose and resist the exploitative practices of big tech companies. When they produce content that enhances public awareness of digital platform datafication practices, and the fact that these practices are poorly constrained, these news media can help to mobilize people to seek ways of combating the harms of mis- and disinformation, thereby contributing to healthier information ecosystems.⁸³

In *Latin America* the rise of alternative news media outlets is attributed partly to the necessity for democratic political communication to address imbalances in information and power. In this region, alternative news media played a role historically in countering dominant transnational communication patterns and cultural imperialism. These outlets often disseminate counter-information and express dissent against the establishment with the goal of facilitating political change. Digital native news sites are operated by professional journalists who generally follow the same professional standards and practices as mainstream media.⁸⁴

The extent of the use of social media for political participation and as a form of alternative news media varies across Latin America.⁸⁵ In the 'Ibero-American' area between 2017 and 2020, three main channels for the dissemination of mis- and disinformation were identified: the legacy news media, open social networks (such as X, Facebook and Instagram) and closed social networks and messaging services (WhatsApp, Telegram and Facebook Groups). Research indicated that closed networks accounted for a large share of mis- and disinformation, while the presence of this content in the legacy news media was much lower.⁸⁶

Alternative news media can also be hyper-partisan and disseminate mis- and disinformation.⁸⁷ In *sub-Saharan Africa*, for example:

Media repression through the enactment of draconian pieces of legislation and the brazen capture of legacy media infrastructures by political and economic elites have been followed by the mushrooming of fake online news sites, faceless social media influencers, pseudonymous social media accounts, and coordinated circulation of false and misleading news information through mostly Twitter, Facebook and WhatsApp.⁸⁸

As battles over the 'truth' intensify in the region, mainstream state-owned media have been branded as 'fake news' outlets because of their biased reporting, while private and independent (alternative) news media are often considered as bearers of 'truth'.⁸⁹

⁸¹ SSD EFF (2023).

⁸² See Section 1, Chapter 2 for a discussion of definitions of news media including 'alternative media'.

⁸³ Reiter & Matthes (2023).

⁸⁴ Harlow (2022).

⁸⁵ Mitchelstein *et al.* (2020).

⁸⁶ Guallar *et al.* (2022), supported in part by the Ministry of Science, Innovation and Universities (Ministerio de Ciencia, Innovación y Universidades), Spain.

⁸⁷ Recuero *et al.* (2022).

⁸⁸ Mare *et al.* (2019, pp. 5–6).

⁸⁹ Mare *et al.* (2019).

Alternative Social Media Platforms: In India, the position of alternative news media in opposition to legacy news media is striking because the latter support the right-wing politics of the ruling Hindu nationalist Bharatiya Janata Party (BJP), instead of the opposition.⁹⁰ However, the BJP-led government also embraces the use of alternative social media platforms in response to the removal of accounts of prominent right-wing leaders from mainstream social media, for violating platform policies.⁹¹

Whether a strong presence of alternative news media contributes to a healthy information ecosystem depends on the context. A distinction needs to be drawn between alt-right sites (intensely engaged in spreading mis- or disinformation) and those practicing journalism with a partisan bias.⁹² A rethinking of the legal foundations of news media is called for in the *United States*. A positive approach to press freedom could create a foundation for news media reform such that government would have an obligation to provide access to high-quality media, because democracy requires this.⁹³

3.3 COMMUNITY COLLABORATIVE STRATEGIES

Community collaborative strategies aim to produce or collect data that is responsive to community needs. They include initiatives by a wide range of communities, from the very local to the municipality. They aim to take a citizen-first, not technology-first, approach to data governance.

3.3.1 Indigenous Communities and the Data Sovereignty Movement

Indigenous communities are establishing alternative data practices through the Data Sovereignty Movement. To maintain control and autonomy over their own data, they are building their own physical infrastructures for telephone and the internet, and developing apps, browsers, streaming platforms and messaging services to serve communities in their own languages. Māori activists in *New Zealand* have used audio data from recordings of people in te reo Māori, the Māori language. They hope to use these digital tools to preserve knowledge of te reo Māori for future generations. Community members spearheading the project have rejected multiple offers from tech companies to incorporate te reo Māori audio data into mainstream apps, for example DuoLingo and Google Translate. Explaining resistance to these offers, Peter-Lucas Jones, one of the leaders of the project, said: ‘our data would be used by the very people that beat that language out of our mouths to sell it back to us as a service... It’s just like taking our land and selling it back to us’.⁹⁴

These and other tactics by Indigenous communities draw attention to how algorithms are used to amplify power imbalances.⁹⁵

3.3.2 Municipal Initiatives

In some ways, municipal-level community strategies for data governance go beyond regional- or national-level efforts. These often emerge, for example, in direct opposition to the imposition of so-called ‘smart city’ initiatives that involve building digital surveillance technologies into the public landscape to collect data for a variety of purposes – from policing to development planning.⁹⁶ ‘Smart city’ initiatives are often driven by multinational technology companies that stand to profit significantly from the embedding of their intellectual property in municipal infrastructures.⁹⁷

⁹⁰ Chadha & Bhat (2022).

⁹¹ Bhat (2021).

⁹² Bennett & Livingston (2018).

⁹³ Pickard (2024).

⁹⁴ See Carroll *et al.* (2019); Dibenedetto (2021); Hao (2022); see also Our Data Indigenous: <https://ourdataindigenous.ca>.

⁹⁵ Chonka *et al.* (2023).

⁹⁶ Galić (2022); Rosol & Blue (2022), supported by the Social Sciences and Humanities Research Council (SSHRC) of Canada; see also Purandare & Parkar (2020).

⁹⁷ Cooke (2020).

A global coalition called Cities for Digital Rights was formed in 2018 with the purpose of ‘promoting and defending digital rights in urban context through city action, to resolve common digital challenges and work towards legal, ethical and operational frameworks to advance human rights in digital environments’.⁹⁸ Many of the coalition’s 60 member cities have introduced new policies and practices designed to shape the way that data is produced and used within the city, and to involve citizens directly in decisions about what digital life in the city will look like.

Barcelona as a case study. This is a city that successfully pivoted from a corporate-centric to citizen-centric ‘smart city’ model.⁹⁹ In 2015, after the election of housing and human rights activist Ada Colau as mayor, the city embarked on a process of re-envisioning policy making about data and digital technology as participatory democracy. Barcelona City Council developed the Barcelona Digital Plan as guidance to implement this new approach.¹⁰⁰ Under this framework, the city has developed a set of policies and practices around data and digital infrastructure.

These include, among other things: opening up the digital architecture of the city by developing open standards and prioritizing open-source technology over proprietary systems; integrating local providers into procurement; treating data as a shared resource (owned and controlled by citizens rather than as a commercial asset) by including provisions to preserve citizen data ownership in city contracts with digital providers; and using technology to foster and facilitate civic participation in municipal policy making.

In 2021, Barcelona introduced its AI Municipal Strategy, which identified four key principles: (1) AI may be used in the generation of automated recommendations, but may not be integrated into decision-making systems; (2) algorithmic models and digital databases should be transparent and auditable; (3) robust accountability and liability regimes apply when the use of AI tools results in harm or loss; and (4) strict procurement clauses that protect municipal control of any private or externally provided AI product.¹⁰¹

Municipal policies to improve digital privacy, limit surveillance technology or place guardrails around the acquisition and use of GenAI are increasingly common in the *United States*. Seattle was one of the first cities to enact a ‘Surveillance Ordinance’ in 2013, which required city departments to submit guidelines for how they planned to use these technologies, and what types of data would be produced.¹⁰² Four years later it was amended to remedy the fact that city departments had failed to include analytic software within the language of the original ordinance. The revised ordinance included provisions for holding community meetings prior to city council approval of departmental surveillance technology (hardware and software) acquisitions.

New York City’s Public Oversight of Surveillance Technology (or POST) Act in 2020 was more narrowly focused on the police department, requiring it to produce impact reports and use policies for its surveillance technologies that include various algorithmic tools.¹⁰³ The POST Act has been criticized as ineffective by many of the groups that advocated for its adoption, largely because it entrenches a bureaucracy within the police department that legitimizes surveillance technology without imposing any real mechanisms for independent enforcement.¹⁰⁴

⁹⁸ Cities for Digital Rights: <https://citiesfordigitalrights.org/thecoalition>.

⁹⁹ Fernandez-Monge *et al.* (2024).

¹⁰⁰ Bria (2018).

¹⁰¹ Ajuntament de Barcelona (2023).

¹⁰² Stevenson (2016).

¹⁰³ NYPD (2024).

¹⁰⁴ Dyson (2023).

In the United States there is also a growing movement to limit specific surveillance technologies at municipal level.¹⁰⁵ For example, 21 municipalities now have ordinances banning or restricting the use of facial recognition technology by certain actors or in certain contexts.¹⁰⁶

3.3.3 Social Entrepreneurs and Community-Controlled Technologies and Data Practices

Digital resistance can involve the creation of technologies that are owned and controlled by local communities, usually for limited purposes specific to the needs of those communities. This often involves social entrepreneurs that partner with local communities to develop new technological applications based on non-extractive data practices. These initiatives challenge the dominance of corporate data governance models, offering alternatives for individuals and communities.

- An example is Alt (Alternativa Laboral Trans), a worker cooperative in *Argentina* owned by trans and non-binary people. It offers digital design and development services using a non-data extractive business model that relies on open-source software, allowing clients to maintain as much control as possible over the digital afterlives of their work.¹⁰⁷ It also provides support for the development of digital tools for the protection and education of its own community members.
- In the *United States*, the National Digital Inclusion Alliance (NDIA) works with local community organizations to promote digital equity, inclusion and literacy, especially in areas underserved by corporate broadband providers. The NDIA's initiatives aim to empower local community actors to take an active role in pursuing digital equity. Through programs like

the National Digital Navigator Corps, it provides personalized tech support and digital literacy training to underserved communities, helping bridge the 'digital divide'.¹⁰⁸

Another example comes from a community in inner city Milwaukee, Wisconsin, which undertook a partnership with university researchers to create their own geographic information system (GIS) that would enable community members to engage politically on questions of greatest concern to them. The aim was to create a database that would, among other things, help them to identify property sales in their neighborhood, absentee landlords who had abandoned properties in the neighborhood, leaving them to deteriorate, and to identify tax delinquency and building code violations.¹⁰⁹ Relying on a combination of public data sets, local data and the knowledge of local community members, the project members built a database with a neighborhood map interface, allowing users to retrieve the data necessary to inform their participation in neighborhood planning processes.¹¹⁰

Alternative principles for data use.

A growing number of communities and organizations is articulating alternative principles around data use and enacting these in practice as they create their own digital tools or resources. The Distributed Artificial Intelligence Research (DAIR) Institute was established to undertake research that benefits 'communities which are typically not served by AI and to create pathways to refuse, interrogate, and reshape AI systems together'.¹¹¹ Other initiatives include a tool using computer vision and satellite imagery to visualize the impacts of spatial apartheid in South Africa,¹¹² machine learning (ML) to analyze the history of racial justice protests

¹⁰⁵ Tate-Mosely (2023).

¹⁰⁶ McConvey (2024).

¹⁰⁷ ALT Cooperative: <https://altcooperativa.com>; Mejias & Couldry (2024).

¹⁰⁸ Menon (2024); NDIA (2023).

¹⁰⁹ Ghose (2001).

¹¹⁰ Ghose (2001).

¹¹¹ DAIR Institute (n.d.).

¹¹² Tsanni (2024).

in the United States,¹¹³ and a wage theft calculator that estimates how many people's wages are lost (stolen) by a surveillance technology called Mentor.¹¹⁴

In the Global Majority World, social entrepreneurs and NGOs are working with local communities to build digital information networks that serve the specific needs of, and are at least to some extent controlled by, communities seeking to avoid participating in data extractivism.

Building community-controlled networks.

Uganda Flying Labs (which is part of a larger humanitarian organization with similar projects in 32 countries) uses drone technology for disaster relief, services planning for refugees and agricultural monitoring and community development. Its data was used to identify evacuation access points during the 2019 Bushika landslides, to coordinate Covid-19 responses in local refugee settlements, and to help coffee farmers respond to extreme and unpredictable weather patterns resulting from climate change. Flying Labs coordinates with local government to implement projects based on the data it collects, but the local government does not own the data. 'Because the use of drones for development-oriented initiatives requires private sector investment, data from the initiatives designed for public goods ultimately is returned to the organizations, company, or entities that funded the request; data are owned by the people who commission research rather than the Flying Labs or the broader community.' This, along with the lack of resources available to establish servers for storing and managing data, means the data that Flying Labs collects cannot be co-opted for other purposes.¹¹⁵

In Cape Town, *South Africa*, a non-profit organization – Violence Prevention Through Urban Upgrading (VPUU) – has developed its own data infrastructure, including a licensed community wireless network serving over 65,000 people for free.¹¹⁶ Called V-NET, it is 'composed of nodes established around community sites', which function as a 'mesh network'. This network was developed to provide internet access to local people and to support the organization's other community development projects, which include educational and social programs, public works projects to improve local infrastructure, and advocacy with the government for better service delivery. VPUU develops its own apps to collect data collection in support of its projects, and trains community members to use those apps. VPUU's CitySpec app allows community members to track the maintenance of public facilities, such as water taps, toilets and streetlights, using the data to analyze community needs; the community then uses that data as evidence to advocate with local government actors to have those needs met.¹¹⁷

Another example is the non-profit organization Majal, founded in 2006 by Bahraini social entrepreneur Esra'a Al Shafei, which operates in the *Middle East* and *North Africa*.

Spaces for safe expression online. Majal develops digital spaces to foster safe and accessible expression, association and communication for underserved and underrepresented communities in the Middle East and North Africa. One of its platforms, CrowdVoice.org, has collected, curated and contextualized crowdsourced data about global protest and social justice movements. The platform serves as a resource for activists and journalists to document, research and communicate about events on the ground. Majal is also a co-founder of the Numun Fund, which aims to 'seed, resource and sustain feminist and women/trans led groups who engage with technology in their activism'.¹¹⁸

¹¹³ Oliver *et al.* (2022).

¹¹⁴ Williams (2023).

¹¹⁵ Horst *et al.* (2024, p. 137), supported in part by the Australian Research Council (ARC).

¹¹⁶ Blake *et al.* (2023).

¹¹⁷ Blake *et al.* (2023); VPUU (2019).

¹¹⁸ Majal (2024a, b); Skalli (2023).

All these initiatives underscore the importance of local control and community-specific solutions in addressing digital inequalities and injustice. Unlike governance measures that merely seek to balance investment priorities or ban certain applications, these initiatives emphasize building local capacity and agency. They demonstrate that empowering communities using tailored digital solutions can effectively address their unique challenges, ensuring that technological advancements do not perpetuate existing inequities, but instead promote inclusive growth and empowerment.

3.3.4 New National-Level Decentralized Data Governance Frameworks

Even when formal institutional policy processes are captured by industry interests or dependencies to varying degrees, communities can engage in organized resistance. Globally, there is great variety and creativity in the resistance strategies that communities have adopted in dissent from both corporate-controlled datafication and from the institutionally controlled political processes of datafication. Most documented examples of organized resistance are small scale and local, but they demonstrate the possibility of, and provide inspiration for, meaningful collective action in contexts where it is difficult or impossible for most people, let alone the vulnerable, to access formal political processes.

New data governance models offer promising alternatives to traditional centralized systems by promoting local ownership, control and benefit-sharing of data. The principles outlined in *India's* Non-Personal Data Governance Framework and the digital citizenship initiatives in *Latin America* provide a solid foundation for developing these models. By implementing such frameworks, communities can ensure that they are the primary beneficiaries of the data they generate, potentially leading to more equitable and sustainable development outcomes. These models not only foster a sense of ownership and empowerment among community members; they also help build trust and cooperation in the increasingly digital global landscape.

Rethinking data governance. India's draft Non-Personal Data Governance Framework is a pioneering effort to rethink data governance focusing on non-personal data generated in rural and urban areas. The framework proposes that data generated by rural *gram panchayats* (village councils) and urban municipalities be owned by local bodies, referred to as data stewards. This is intended to ensure that the benefits of data exploitation are democratized, and local communities have a say in how their data is used. Data stewards or trustees serve as data guardians, responsible for managing and regulating access to this data. They are tasked with ensuring data privacy, securing data rights and fostering a transparent environment where community members are informed and engaged in decision-making related to their data.¹¹⁹

These kinds of data governance frameworks can support a variety of community-based products and services:¹²⁰

- *Community data hubs* are localized data centers where community data is stored, processed and managed, with oversight from local data stewards. These hubs could serve as centers for innovation and learning, offering training programs on data literacy and data rights.
- *Decentralized data infrastructures* use blockchain or other decentralized technologies such as IPFS (InterPlanetary File System), decentralized identity systems and smart contracts to ensure that data transactions are secure, transparent and accountable to the local community.
- *Local data analytics services* are developed within the community to analyze local data and provide insights that directly benefit the community. These might include agricultural advisories in rural areas using local climate and

¹¹⁹ Bailey *et al.* (2020); Jindal & Nigam (2020); see also data trusts in the African context (Olorunju & Adams, 2024).

¹²⁰ Kumar *et al.* (2023); Lanier & Weyl (2018); Micheli *et al.* (2020); Singh (2020); Verdegem (2021).

soil data, or urban planning tools in cities using mobility data to improve public transportation systems.

- *Data lockers* and intermediaries enable people to control their own data. However, this presumes that people have the time and knowledge to make judgments about when to release their data and to whom.
- *Data cooperatives* governed by community rights frameworks so, for example, women, farmers and others can retain control of their data and the rewards that flow from its use.
- *Public data infrastructures* are being developed in India in sectors such as commerce, finance, health, education and agriculture, and these may become more common globally.

Commons-based approaches to data governance offer an alternative paradigm to proprietary models. These typically leverage cloud-based software platforms with open governance structures, allowing data to be managed, analyzed, accessed and shared within a community.

Commons-based data governance.

Collaborative models are rooted in principles of open access and collective benefit, drawing on practices in free and open-source software (FOSS) communities. These have limitations when data is treated as a common-pool resource (due to its characteristics of non-excludable and non-rivalrous characteristics, and challenges in excluding beneficiaries). In the context of alternatives to mainstream governance models, collective management of data to support the political claims of communities can mean that communities classify data as a commons. This means data is stored and

actively managed and analyzed using shared tools, significantly reducing redundancy and enhancing the quality of data analysis. A 'data commons' democratizes data by breaking down barriers that hinder collaborative research and innovation. Commons-based models offer a promising alternative to prevailing models of data management as well as AI development. Achieving these outcomes requires careful implementation of governance structures, privacy protections and a technical infrastructure.¹²¹

All these approaches can facilitate data sharing within and between communities, promoting collaborative projects.¹²² They may be resource-intensive and some have limited scalability. Local systems require periodic updates and hardware and software maintenance, which is often costly. They may also lag in adopting the latest global technical advances due to their isolation, and the focus may be on local data processing. In Global Majority World countries, research is examining frameworks for the collective realization of the social value of data, 'meaningful data transparency' in data access in the case of African stakeholders, and other efforts to stop large technology companies from being the 'privileged providers of *social solutions*'.¹²³ A commons approach is proposed as a way of dismantling the concentration of power in the AI industry sector, extending beyond a data commons to include the infrastructure of computing power to create a 'communal utility', and ambitions that clearly involve a rethinking of ownership.¹²⁴

3.3.5 Civil Society Organizations, Researchers and Philanthropic Organizations

Civil society organizations can work to establish systems of social support for communities harmed by tech dominance and datafication. They often collaborate with researchers to mitigate harms to

¹²¹ Berdou (2011); Birkinbine (2018); Dalle & David (2007); Powell (2015); Zygmuntowski *et al.* (2021). It should be noted that commons-based strategies can become conflictual when they intersect with commercial and state efforts to appropriate their resources.

¹²² Menon (2024).

¹²³ See Magalhães & Couldry (2021, p. 354; emphasis in original); see also Adel *et al.* (2023); Gurumurthy & Chami (2022); Nyalety *et al.* (2019); Omar (2023); Page *et al.* (2023); Vayadande *et al.* (2024).

¹²⁴ Verdegem (2022).

individuals and communities, and remove barriers to organizing for those who are most negatively impacted by datafication. They help to build literacies in local communities and the capacity of local communities to use data on their own terms or to resist its use against their interests.¹²⁵ They undertake critical research to expose harms of data-intensive tools and infrastructure, and identify strategies for data governance that is democratic in practice, not just in stated policy. They add knowledge about the political economy of data, enhancing public awareness of the epistemic and material harms of poorly constrained datafication.

Researchers and civil society organizations that focus on data justice and engage in data activism are critical of technology-centered approaches.¹²⁶ Their work can help to amplify the reach of social movements supporting democracy and people's capacities to control the uses of data and technology, including AI, in line with human rights commitments and reducing inequalities.¹²⁷ With funding from philanthropic organizations, by crowdsourcing funds or drawing on (scarce) university funding, they contribute to local initiatives (and sometimes global initiatives) to redistribute power over data, and over decision-making about data, away from the big tech companies. Research in this critical tradition identifies how concentrations of power in the hands of large companies and states leads to intrusive datafication and surveillance.¹²⁸ It focuses on 'the absences, the silences and the forgotten and ignored people and regions of the world',¹²⁹ contributing to resistance to corporate datafication, as outlined in Table 8.1, by developing, facilitating and supporting participatory action research in collaboration with and within local communities.

'Good' or positive outcomes occur ideally when initiatives are established within and by

communities, and when outcomes are not assumed to be driven by technology but by actor choices about the design and operation of technology. This notion of technology for the 'public good' differs from the way 'digital public goods' are often discussed by United Nations agencies and others. These discussions tend to assume that 'the good' is embodied in technology and in data. In such discussions it follows that technology and data only need to be made available to those without affordable access for them to reap the benefits.¹³⁰ For example, a digital public goods alliance of governments and the private sector is working to deliver 'digital public goods' for the Global Majority World. Its mission is to fight against mis- and disinformation, and it promotes creative uses of technologies and data. Some of its projects are open source, but few seem to push for the step-change in data governance frameworks that would empower local communities to control and own their data; they do not fundamentally question that big tech's datafication practices lead to harms, including discrimination.¹³¹

4 Chapter Summary

This chapter has demonstrated that commercial datafication supported by AI systems (data aggregation and ML technologies) disadvantage and discriminate among people in the data economy by sustaining comprehensive surveillance to enable computerized data production and services. These surveillance practices are designed to monopolize data resources. The monopolization of information (i.e., organized as usable insight or knowledge), as practiced, converts data into private assets. Big tech business models incentivize turning a blind eye

¹²⁵ See Chapter 5 for a discussion of literacy.

¹²⁶ Crawford *et al.* (2014); Dencik *et al.* (2016); Hepp *et al.* (2022); Milan & van der Velden (2016).

¹²⁷ Cammaerts (2018, 2024); Dencik & Leistert (2015); Ó Siochru *et al.* (2024); Timcke & Hlomaní (2024).

¹²⁸ On surveillance using biometrics, see Munoriyarwa & Mare (2022); on the use of facial recognition technologies in Brazil, see Ramiro & Cruz (2023); on surveillance in African countries and the use of technologies exported from the Global North, see Sheombar & Skelton (2023); on the 'new aesthetics of surveillance' using digital images and the systematic collection of data, see Beiguelman (2021); on the impact of AI systems-enabled surveillance and data collection on migrants and refugees, see Napolitano (2023); and for the use of surveillance or 'smart spying' by the United States, see Moran *et al.* (2023); in the European Union, see Calderaro & Blumfelde (2022), supported by the Economic and Social Research Council (ESRC), UK.

¹²⁹ Gillwald & Wavre (2024, p. 34); RIA (2023a).

¹³⁰ UN (2020). Digital public goods in this context refer to open-source software, open data, open AI models, open standards and open content.

¹³¹ DPGA (2023).

to mis- and disinformation because this content is key to attracting attention and traffic to platform services, which boosts their financial viability and profits.

Corporate incentives, strategies and practices involved in designing, developing, selling and controlling data are at the heart of information ecosystems. These lead to epistemic injustice, the privileging of information and knowledge that are neither representative nor inclusive. Individual and collective dependencies and inequities resulting from datafication are being experienced around the world. They are manifested in the form of biased and discriminatory decisions in the treatment of people by gender, race, ethnicity, disability, class (caste) and language community. To progress towards healthy information ecosystems, capacities for thinking critically about how to govern massive amounts of digitized data need to be strengthened.

The synthesis of research in this chapter shows that:

- Biased outputs of AI systems are often the consequence of biases in the data on which they are trained. This leads to distortions and unfair discrimination, inflicting harm by causing unfavorable outcomes for groups by gender, race, ethnicity, disability, class (caste) and language community.
- Improving data diversity by enforcing transparency and conducting regular bias audits and algorithmic reviews is essential because bureaucratic oversight mechanisms mandated by state-led governance have had limited impact on the power wielded by big tech companies. These audits should be facilitated by third-party auditors, using fairness tools to adjust AI models to ensure they are free of known biases.
- Individuals and community groups are developing strategies to resist the extractive features of the data economy. There are strong pressures from within civil society to treat data governance as a lever for restructuring data markets, to protect against infringements of

human rights and to tackle concentrations of power and wealth that jeopardize democracy.

- Confronting data extractivism through resistance strategies requires scaling up digital self-defense training. Other strategies include the development of public interest alternative news media, promoting community collaborative strategies with Indigenous communities and municipalities, working with social entrepreneurs to develop community-controlled technologies and data practices, and decentralized data governance frameworks. These require working with civil society organizations, researchers and philanthropic organizations to counter big tech datafication practices and to achieve data justice.

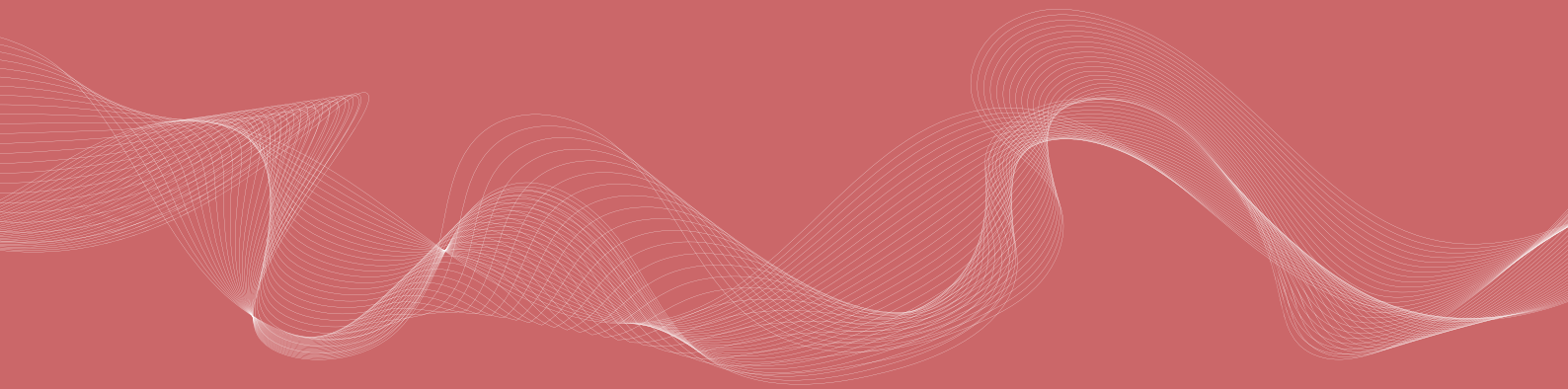
Research is needed:

- To advance work on decolonizing research so that epistemic knowledge about and experiences of the data economy in the Global Majority World can be understood and inform data governance policy and practice.
- To examine the impacts of data production and processing in people's daily lives, focusing on discriminatory outcomes by gender, race, ethnicity, disability, class (caste) or language community; this means extending research to capture instances of these outcomes in countries around the world.
- To expose how dependencies created by the power of big tech companies in other sectors, for example healthcare, education, transportation and the news media, pose significant risks to democracy when sectors become dependent on the data-intensive products and infrastructures developed and sold by big tech companies.
- To investigate how people are imagining resistance strategies to challenge biased algorithmic systems and injustices associated with data governance frameworks, and to systematically identify knowledge about practices and local solutions that may be sustainably scaled up.

- To map new digital divides that are emerging with the spread of AI systems (data aggregation and ML technologies), and to investigate how to prevent replicating and further entrenching problems with data-intensive economies that are present in the Global North and the Global Majority World.
- To understand how a paradigm shift can be achieved such that the Global Majority World is not positioned as a passive recipient of Western ideas about how to govern data, but as an equal stakeholder in dialogue about information ecosystems governance.

CHAPTER 9

CONCLUSION: INFORMATION ECOSYSTEMS AND TROUBLED DEMOCRACY



The map is generated by the OID on the basis of the chapter's text using GarganText – developed by the CNRS Institute of Complex Systems. Starting from a co-occurrence matrix generated from chapter's text, GarganText forms a network where words are connected if they are likely to occur together. Clustering is conducted based on the Louvain community detection method, and the visualization is generated using the Force Atlas 2 algorithm.

i

'Information and freedom are indivisible'
(Kofi Annan, 1997).¹

1 Introduction

This report is a critical analysis of state-of-the-art research in the Global North and the Global Majority World that informs us about the interdependent relationships between the cultural, social, political, economic and technological components of information ecosystems. The analysis in the preceding chapters has focused on what interdependence means for the integrity of information and for informed democratic participation in the public sphere.² The report addresses three thematic areas with a cross-cutting theme of mis- and disinformation: **media, politics and trust; artificial intelligence, information ecosystems and democracy; and data governance and democracy.**

This research assessment was completed in October 2024, just after the United Nations published a *Global Digital Compact*, highlighting the key challenges and declaring its intent to address them:

We must urgently counter and address all forms of violence, including sexual and gender-based violence, which occurs through or is amplified by the use of technology, all forms of hate speech and discrimination, misinformation and disinformation, cyberbullying and child sexual exploitation and abuse. We will establish and maintain robust risk mitigation and redress measures that also protect privacy and freedom of expression ... [protecting] the rights of the child in the digital space, in line with international human rights law, including the Convention on the Rights of the Child.³

From the outset, when work started on our research assessment, we recognized that the design and

development, as well as the beneficial and harmful uses of digital technologies, are not simply driven by technological change; they are the result of human decisions and human action. These depend on power relationships embedded in institutions and technologies and how these change over time as a result of the actions of governments, companies, civil society organizations and individuals. For this reason, we understand information ecosystems as a 'system of people, practices, values, and technologies in a particular environment', embedding the public sphere within two layers of the ecosystem: a network infrastructure (hardware and software) layer and a service applications layer.⁴

These actors establish the norms and rules that govern how information ecosystems develop. Rules and norms matter because they affect whether internationally agreed human rights are protected, and whether a public sphere for informed democratic debate will thrive. 'Information and freedom' are indivisible, as former Secretary-General of the United Nations, Kofi Annan said, but information that is complicit in eroding individual rights and collective societal interests in fairness and justice can amount to illegal or harmful mis- and disinformation. When mis- and disinformation are pervasive in the public sphere, we treat this as both a symptom of complex changes in society and as an important amplifier of these changes.

The research assessment was based mainly on academic publications and supplemented by reports and other materials from different disciplines and regions (1,664 citations selected from our bibliographic database, with 3,095 entries screened before inclusion). In view of the speed of change and the currency of debates about

¹ Annan (1997, n.p.), then United Nations Secretary-General.

² See Section 3, Chapter 1 for definitions of concepts including information ecosystems, information integrity and the public sphere, and see Appendix: Methodology for the rationale for their use.

³ UN (2024b, para. 30).

⁴ Modified from Nardi & O'Day (1999, p. 49). We acknowledge that there are many ways of defining these ecosystems including those that assume a rhizomatic systems dynamic among the components, see Radsch (2023e).

intersections between corporate and political power and technology and human beings in today's information ecosystems, 'other materials' includes preprints, conference papers and online sources such as blogs, newspapers and magazines. Research was included principally from the fields of computer science, engineering, law, media and communications, political economy, political science, psychology and sociology. It included works concerned with identifying the impacts of digital technology or mis- and disinformation on society or individuals, and research considering the contexts in which information and technology have come to be seen as problematic, why, and for whom. Multiple research design and methods (qualitative, quantitative and experimental, and mixed methods) were included. The emphasis was on recent published works, rather than an attempt to provide an in-depth account of the history of research in the areas covered.⁵

The decisions and processes involved in undertaking this research assessment, including the steps taken to include work in the Global Majority World (22.5% of citations in this report; Global North 65.5%, Global 12%), selection criteria and quality checks are described in Appendix: Methodology. This global research assessment is not intended to prescribe specific actions for policy makers; rather, it showcases what we can learn from landmark research on the often intractable challenges posed by the rapid changes in information and communication spaces.

Sections 2 and 3 of this concluding chapter discuss central themes that emerged across the chapters of this report. Section 2 highlights insights arising from the analysis of research findings, while Section 3 focuses on the key characteristics of state-of-the-art research, again focusing on insights drawn from a cross-cutting review of the preceding chapters. Section 4 provides a chapter-by-chapter summary, highlighting the core research questions and key findings (readers should go to the individual chapter summaries for a full account of findings and priorities for future research). Section 5 sets out the limitations of the report, and Section 6 concludes

with actionable insights that point to what can and could be done to address the 'information crisis'.

2 Principal Thematic Insights

Here we highlight selected themes and conceptual arguments that appear across the topics and issues examined in the chapters of this report (see Figure 9.1). Research communities involved in undertaking studies on questions arising in the thematic areas of interest in this report tend, on the one hand, to welcome the rapid pace of technological innovation and deployment, with a view that any harms will be mitigated, and prioritize getting these technologies to market as quickly as possible for the benefit of humanity. On the other hand, we also draw on insights from research communities that signal the need for caution. In this case, while the many benefits of new technologies (including AI systems) may be recognized, how new technologies become embedded in society and in individual lives is a matter of choice – choice that occurs in a world with unequal power relations.

Figure 9.1
Principal thematic insights



Source: Authors of the report.

Research of this kind typically calls either for greater efforts to introduce mandatory governance for big tech companies or observes that benefits are not being equitably distributed. In this case, the research highlights the need for deeper questioning of the logics of corporate business models and the priorities of those who govern in ways that do

⁵ Footnote 1 in Chapters 2 to 8 provides citations for readers seeking background information.

not uphold universal human rights commitments. In brief, the principal thematic insights emerging from this report range from welcoming recent developments in information ecosystems, to skepticism, to outrage.

2.1 HUMAN RIGHTS

Discussions of human rights commitments appear frequently in research on the news media, AI systems and data governance (including governing mis- and disinformation). This includes, but is not restricted to, calls for legislation to protect human rights or to legal interpretations of existing law. Our critical analysis indicates broad agreement that states have a duty to protect human rights and fundamental freedoms, which includes a negative obligation not to violate rights and a positive obligation to protect human rights.

We found no disagreement that internationally protected human rights and fundamental freedoms are pertinent to information ecosystems. In most instances, it is recognized that states need to ensure that their obligations to respect, protect and implement rights are responsive to the challenges posed by information ecosystem actors and instruments. However, it is insisted in the literature that it is essential to differentiate between the normative goals and principles articulated at a global level, and how these are translated over time into practice at local, country and regional levels. Even if international human rights commitments bind signatory states, these are interpreted in different ways, and implementation may not be consistent with normative expectations.

The need to protect media freedoms, freedom of expression and to avoid suppression of voices for political reasons is consistently emphasized. While news media freedom has never been absolute and journalism privileges vary, human rights principles should guide normative expectations for the role of the media, even when there are deviations in practice.

There is substantial evidence that the use of AI systems in content governance can lead to human rights violations. It is well documented that automated content governance and algorithmic

decisions can impact negatively on democratic decision-making processes when these systems determine the conditions under which content is seen and with whom it is shared. More generally emerging technologies, such as generative AI (GenAI), challenge both individual rights and rights to democratic participation.

In the literature on data extraction, processing and use, there are several recurrent themes. One is that risk mitigation strategies and practices – voluntary or mandated by legislation – are the preferred means to protect individual privacy, and that corporate appropriation of data generated by people's online interaction is key to prosperous data economies. Another is a questioning of the legitimacy of big tech company data extraction practices and the monetization of data for profit, based on evidence that this leads to unacceptable outcomes, including discrimination and inequalities. In this view, human rights protections are insufficiently robust and the commercial datafication model needs to be reimaged and resisted in the collective interest.

Across the issues addressed in this report (media, politics and trust; AI and democracy; and data governance), there is a clear need for research on how international human rights law is interpreted and applied at regional and country level, and whether commitments to protect fundamental rights are actually being met.

2.2 CONTESTING DATA MONETIZATION

Asymmetrical power relationships and their consequences for strategies and practices of data monetization is a consistent theme in research on changes in the news media industry, AI systems development and use, as well as in research on the role of data in economies. Research repeatedly draws attention to why and how these relationships can lead to disadvantage and discrimination, and the need to acknowledge that these problems arise on both the infrastructure and service applications layers of information ecosystems.

For example, on the infrastructure layer, network neutrality policies and ‘zero-rating’ contracts are impacting who is connected and who can be disconnected during elections or political unrest, who can access various sources of information, and whether the information ecosystem favors informed participation in the public sphere. This report focused, to a limited extent, on the underlying infrastructure, but it is clear that there needs to be research on the fragmentation or ‘balkanization’ of the internet, measures to strengthen digital sovereignty and the ambitions of big tech firms and infrastructure service providers alongside research on information integrity problems on the applications layer of information ecosystems. Understanding developments on the infrastructure layer should inform assessments of the health of information ecosystems and acknowledge that the implementation of network neutrality policies and restrictive contracts on data and information access have markedly different effects at different national (and local) contexts.

The news media industry is consistently shown to be influenced by data monetization strategies and AI systems and algorithm developments. These create incentives for legacy news media concentration, destabilize news organizations financially, and lead to closures, especially of smaller local news outlets. Evidence confirms that power asymmetries are at the core of struggles between the news media industry and the big tech company platforms. Power asymmetries are similarly an issue when governments, political parties and other actors manipulate information using datafication (personalization) techniques during critical election periods, and mis- and disinformation are weaponized by both domestic and foreign actors.

Our analysis highlights the need for an insight into whether technical competencies are in place to enforce measures to combat harms, especially in times of conflict, but also whether such measures are consistent with a diverse public sphere.

Asymmetrical power is also visible in research on the monopolization strategies of big tech

developers of AI systems and their data practices. In the critical literature, these strategies are found to be misaligned with individual and collective interests, and facilitate the production of mis- and disinformation. Efforts to change these strategies require collective civil society mobilization beyond the capacities of either states or individuals. A repeating theme is that policies that favor the data dependence of private and public organizations as well as individuals are preempting the development of information ecosystems. These include meaningful political deliberation on issues such as rights to data ownership, what role data should have in the private and public sectors, and what contexts require the minimization or prohibition of data production. A further prominent theme is the need for systematic research on data activism initiatives that aim to reimagine ways of restructuring data markets to diffuse concentrations of power that jeopardize democracy.

Counter-power strategies would clearly benefit from research aimed at exposing how big tech business models make them attractive targets for mis- and disinformation campaigns, and how digital platforms abandon or arbitrarily change content self-regulatory measures, lay off staff, weaken privacy policies or impose limits on fact-checking.

2.3 EXCLUSION AND INEQUITABLE INCLUSION

There is a common neglect of the distinctive characteristics of information ecosystems at local, national and regional levels, and especially of differences between the Global North and Global Majority World (and within the Global Majority World). When research considers the impacts of mis- and disinformation, too often the implicit assumption is that these findings can be broadly generalized. Even when this is not the case, as in larger-scale comparative studies that capture impacts mainly at the country level, local and rural/urban experiences are left out of the analysis.

When it comes to assessing the characteristics of trust in news media or in AI systems products and the consequences of how they are infiltrating

people's lives, much of the research we assessed does not consider that internet connectivity is absent for many in the Global Majority World, and meaningful connectivity (affordability, skills and outcomes) is unevenly distributed. News media systems are themselves subject to different ownership, and regulatory regimes and content governance measures can often suppress voices that are critical of state authorities. AI-enabled algorithms promote/demote content in different ways depending on country conditions and political influence. AI systems are deployed in ways that impact communities of color, women, religious minorities and LGBTQ+ people in harmful, yet different, ways. These and other conditions mean that there are substantial differences in how people in low- and middle-income countries, as compared to high-income countries, experience information ecosystems.

In some literature exclusion and inequitable inclusion are discussed with warnings that failure to take difference into account is a recipe for replicating and exacerbating inequalities and injustices. These warnings are present in concerns about individuals, communities and countries becoming dependent on digital infrastructures and algorithmic products produced by big tech companies in the Global North. There is also growing discussion about an 'AI divide' (a growing disparity between those who can access and effectively leverage AI systems and those who cannot). A key theme in our analysis is that homogeneous approaches to governing AI systems and tackling mis- and disinformation are misguided, but there was very little evidence of research on AI system investment strategies being developed that aim to foster international solidarity and inclusive participation.

Inequitable inclusion is also visible in debates in the literature about whether AI systems can be free of bias. Some research insists that they can, but the most prevalent view is that no algorithm or training data set can be free of bias. No content moderation or content curation system can be neutral – there is always the potential for these systems to be used to pursue politicized agendas. As a consequence, outputs of large language models (LLMs) cannot be

expected to be fully representative or inclusive on equitable terms, and AI algorithms deployed in the media industry will reflect biases as the result of decisions taken about their design and operation. Some form of epistemic injustice – the privileging of particular kinds of information and knowledge – is always going to be present. The question is how best to counter it.

Research points to policies for media freedom, responsible development of AI systems and novel approaches to data governance, but there is a clear need for more work to track the ongoing experiences of Global Majority World countries as they seek to fashion their information ecosystems in ways that are both just and responsive to their conditions.

Addressing injustices is shown in the literature to require critical thinking about how to govern news media, AI systems and data to counter exclusions and inequitable inclusions. On the policy level, including the *Global Digital Compact*, there are ambitions to tackle exclusions from, and inequitable inclusion in, information ecosystems, and to support measures aimed at enhancing information integrity. In the academic literature there is much discussion of the problems, but little evidence of systematic practical steps to bring about a paradigm shift that would ensure the Global Majority World is not a passive recipient of Eurocentric/Western ideas.

There is a clear need to reduce barriers to participation by people in the Global Majority World in all facets of decisions about information ecosystems (including how to treat mis- and disinformation). It is all the more critical to remove these obstacles since they affect the development of AI systems standards and practices. This means devising practices and resourcing them to find creative approaches that ensure that elite Global North knowledge is not the unquestioned guide to governing information ecosystems and the public sphere.

2.4 TRANSPARENCY AND ACCOUNTABILITY

The concepts of transparency and accountability are pervasively discussed in the literature on governing the news media, AI systems and data, and these appear as being relevant to research on media and information literacy (MIL) and AI literacy.

2.4.1 Governance

A consistent theme is that information ecosystems governance, on the one hand, is too permissive and, on the other, not permissive enough. This varies by topic and by context (whether governance involves democratic or autocratic states). Where governance is found to be too permissive, for example in permitting big tech business to foster the amplification of mis- and disinformation, this is because they are found to privilege their economic self-interest without sufficiently strong rights protections. In short, governance is not strong enough to hold big tech companies accountable, and a lack of transparency in corporate data collection is allowed to persist along with targeted advertising or misuse of data for political gain.

When governance arrangements are found not to be permissive enough, this is typically because states are found to be exerting undue pressure, leading to the suppression of voices. There are considerable differences in views in the literature about how accountability of state and corporate actors is best achieved, that is, through discretionary or mandatory measures. These differences depend on which values receive priority.

Our assessment indicates that governance initiatives are needed to tackle the monopolistic power of big tech companies when it is found to unfairly reduce competition and, in some jurisdictions, to lead to harms to privacy. Governance also needs to be strengthened around data collection and to reinforce measures to control stakeholders involved in sharing and selling data. These views are common in both the Global North and Global Majority World, but in the latter, there are concerns about the feasibility of holding distant actors to account.

Another recurrent theme in the literature is that governance models – for example the European Union’s Media Freedom Act of 2024, General Data Protection Regulation (GDPR) or its Artificial Intelligence (AI) Act of 2024 – should not be treated as a panacea for all threats and harms linked to mis- and disinformation. Overall, there is little clarity about the most crucial intervention points where governance can have positive impacts. This means that there is no shared understanding of the appropriate balance between the imperatives of economic growth, innovation and human rights protections when it comes to designing governance to combat mis- and disinformation; this is partly because of concerns about the risks of regulatory overreach by governments, particularly by authoritarian governments.

In the case of measures to promote AI systems transparency and ethical practice in the newsroom and other settings, it is often unclear who is held accountable for harmful outcomes. There are numerous calls in the literature for regular AI systems audits, but less often about who might perform these audits and how they might accomplish them. In the case of AI systems and content governance, there is much evidence of calls for a greater focus on explainability and the development of accountability best practice, but research indicates that the public is unsure about who is responsible for protecting their rights.

To hold the big tech companies and governments to account, accurate information is essential in the hands of a wide range of stakeholders. Those whose evidence questions current practice should not be criminalized or marginalized for holding opposing views or for exposing how their interests are not protected.

2.4.2 Literacy Issues

The challenges created by mis- and disinformation for news media, AI systems and data governance direct attention to promoting enhanced media and information literacy (MIL) (sometimes called digital literacy) as well as AI literacy (data literacy, algorithmic literacy) for individuals (designers of

technology systems, children and adult users of online systems). Much of the literature is concerned with curricula, training and funding, but literacy issues also make an appearance beyond studies of this kind.

Research is clear that MIL and AI literacy policies and initiatives should focus on more than technical skills. Research emphasizes the need to hone critical literacy skills, and for attention to how these skills can be taught effectively to children and adults. Some evidence indicates that those with critical literacy training are less susceptible to mis- and disinformation, although most research only examines the short-term impacts of training and finds a lack of resources, particularly in the Global Majority World. Studies of AI literacy training indicate its necessity at all stages of AI systems development and deployment. It is also clear that literacy initiatives cannot be seen as a solution to all information ecosystems problems, including declining trust in information in the public sphere.

MIL and AI literacy should never be presented as a stand-alone project aimed at keeping individuals safe from mis- and disinformation – it must be accompanied by state-led (as appropriate, in view of human rights protections) and individual- or community-led responses to the information crisis.

Less discussed, but making an appearance in the literature, is that literacy training is also important on a societal level and not just an individual level. Educating the public about the complex issues facing information ecosystems is paramount. An informed public is more capable of demanding accountability from big tech companies and states to ensure that changes in information ecosystems respect human rights. They will be better equipped to insist on the transparency (as far as possible) of algorithmic systems, on human oversight of algorithmic decisions about their lives, and generally, to participate in the public sphere in an informed way. Stronger MIL and AI literacy among policy makers is also essential if they are to devise effective accountability frameworks, monitor and

enforce them. We found little evidence on the kinds of research evidence relied on by policy makers in deciding how to govern information ecosystems, counter mis- and disinformation and strengthen democracy.

There is little systematic evidence of experience over time on literacy initiatives on a global basis, although the evidence we do have suggests that it can make a positive contribution to individuals' efforts to keep themselves safe online, and to make sense of the information they encounter if training is well resourced.

3 State-of-the-Art Research and Future Directions

This section provides a critical assessment of state-of-the-art research focusing on consistent themes across the chapters in this report. These themes are related to the Eurocentric/Western bias of research, to the conceptual framing of research, to diverse research design and methods, constraints on researchers' access to data and the independence of research activity (see Figure 9.2).

Figure 9.2
State-of-the-art research assessment



Source: Authors of the report.

3.1 A EUROCENTRIC/WESTERN RESEARCH BIAS

A Eurocentric/Western bias towards research in and on the Global North is as disturbing as it is unhelpful when the aim is to understand the interdependence and power asymmetries of the components of information ecosystems in global contexts. The problems created by mis- and disinformation and approaches to mitigating harms are studied disproportionately in the United States and other Western countries. Large-scale surveys include countries in the Global Majority World, but only some of this work is peer-reviewed, and much of it appears in reports (only some of which are peer-reviewed). The main aim of this research assessment was to examine interactions on a 'system' or institutional level, so we did not include 'case studies' or 'use cases'. If we had included in-depth sectoral or local profiles, there would have been case studies to draw on in regard to news media, although this would have been less likely for uses of AI systems and the challenges of data governance as experienced by civil society groups and activists as well as individual online users beyond the Global North.

Research on which companies – small and large – are involved in datafication processes that yield discriminatory outcomes only focuses on a few companies and does not extend to in-depth assessment of experiences around the world. Longitudinal research on what publics believe should be done about illegal and harmful mis- and disinformation is scarce, and coverage of all countries is missing. We have little insight into which civil society actors are represented in deliberations about how to govern information ecosystems at all levels (local, national, regional and global), and the reasons that they become excluded from these deliberations.

The Eurocentric/Western bias of research in all the areas examined in this study needs to be addressed if the views of individuals and organizations in the Global Majority World that work on mis- and disinformation issues are to inform policy in the Global Majority World and debates at international level. A high priority is to work

towards decolonizing research and the advice provided to governments and other organizations.

3.2 CONCEPTUAL FRAMINGS

The use of multiple definitions of concepts across research fields is striking. There is some consistency in the naming of objects of interest and in the way concepts are defined in policy documents, for example information ecosystem, information integrity, mis- and disinformation and 'AI'. These definitions are articulated at an abstract level, and our analysis indicates that their meaning differs in various regions/countries. Despite a shift towards the adoption of several metaphors in recent years, studies emphasize different components of information ecosystems, interpret illegal or harmful information in very different ways, and take different positions on what information should be amplified or suppressed, and whether the focus should be on the public sphere.

In the research community there is debate about whether our object of interest – the information ecosystem – is the priority, or whether the focus should be on the public sphere. There are differences on whether 'information integrity' is too open to varying interpretations on what is good or 'polluting' information. Some prefer to refer to the 'public worthiness' of information to stress informed public discourse and issues such as visibility, access, reflexivity, mediation, influence and legitimacy.

There are few signs of efforts in recent research to conceptualize issues of mis- and disinformation and information integrity in a way that acknowledges lessons from history. Earlier propaganda research is rarely mentioned apart from in studies of trust in news media, where there are some exceptions. In some research 'information society' or 'knowledge society' issues are prominent, although it is not always clear how these concepts differ from what is explored in the case of 'information ecosystems'. There is slippage between how the words 'data', 'information' and 'knowledge' are conceptualized, and lessons from research on the complexity of information environments rarely appear to inform

studies of information ecosystems, except in research that draws on complexity theories of adaptive systems.

Research focuses on different components of information ecosystems without being clear about what is included or excluded. For example, numerous definitions of ‘news’ appear with a recent strong emphasis on online news that often neglects the role of legacy news media; and research on the role of news media in the public sphere is inordinately focused on news content produced by professional journalists. ‘AI’ is used – misleadingly – across much of the academic literature that focuses on governance issues and it is used as a generic category of digital systems. This is misleading when the task is to respond to specific risks. In contrast, other research is very specific about the object of study, for example LLMs, but takes little account of the social factors that influence system designs and implementations. Conceptualizations of MIL and AI literacy also differ, and there are no standardized definitions across regions.

Fragmentation of disciplines is common to all academic fields. There are persistent calls for holistic approaches to bridge between the humanities, social sciences and sciences, and to capture the whole lifecycle of mis- or disinformation. Common definitions are important for large-scale, comparative studies, but definitional variety is needed to capture different experiences.

In addition to efforts to find common conceptual ground and to be clear about how concepts are defined, efforts to understand how mis- and disinformation are entangled with democracy would benefit from joined-up research with the fields of securitization and the socio-economics of online labor markets. These areas that are not covered in depth in this report, and are rarely cross-referenced in the materials cited in this report.

3.3 RESEARCH DESIGN AND METHODS

In this report we emphasize that a robust view of mis- and disinformation requires analysis of the

complexity of information ecosystems, and this applies to all the objects of interest – the news media industry, AI systems and data governance.

Many studies aim to detect a causal links between mis- and disinformation, changes in attitudes and behaviors and political polarization. Much of this work is informed by theories of media effects, and is undertaken in experimental or quasi-experimental settings or based on respondent self-reporting. The research often points to this kind of information as the ‘cause’ of political polarization. Other studies point to political knowledge, and whether people belong to homogeneous social, political, cultural and economic groups, as the ‘causes’ of social discord and distrust that give rise to polarized public opinion. Similarly, some research points to AI systems and algorithms as the ‘causes’ of changes in attitudes and behavior, and calls for risk mitigation measures. Other studies attribute the causes of instability or conflict to power asymmetries that allow commercial datafication systems to flourish. Studies that find that mis- and disinformation is the cause of filter bubbles and echo chambers are not always sensitive to the conditions in democratic and autocratic regimes.

Fewer studies examine reciprocal relationships between components of information ecosystems (news media, AI systems and datafication processes). Much research focuses on information itself and its impacts, neglecting socio-economic, political and cultural conditions that give rise to it. Also often neglected is the fact that mis- and disinformation are produced and circulated *outside social media*.

Research designs aimed at identifying causal effects of mis- and disinformation on individual attitudes and behaviors needs to be complemented by multidimensional research, on both individual and societal harms, and on the factors in society that give rise to this information.

Much of the research on countering mis- and disinformation is undertaken in experimental or quasi-experimental settings or based on survey respondent reports, and relies on quantitative

evidence and predictive models. It focuses on the effectiveness of technical tools in providing countermeasures more than on complex factors that give rise to this information. For example, content governance methods such as fact-checking are shown to have effects on people's responses to mis- and disinformation, but most studies are one-off, and unable to account for techniques and practices that change over time.

Qualitative (or mixed-methods) research drawing on interviews, focus groups, storytelling, etc., and qualitative data analysis techniques (e.g., thematic, discourse, qualitative content, document analysis) are needed to elicit a deeper insight into complex changes in the public sphere and in the components of information ecosystems, on both the service applications and infrastructure layers.

Qualitative methods can elicit insight into how power disparities – explicit and hidden – influence choices about the design and deployment of digital technologies and the agency of individuals and groups that engage with these technologies and with digital content. For example, qualitative methods can help to reveal why people value online filter bubbles. Typically treated as having a negative impact on democracy, self-imposed filter bubbles are sometimes valued when they provide a safe space to marginalized groups to express opinions and avoid political or social repression. Qualitative research on AI-driven mis- and disinformation campaigns can provide fine-grained insight into how these processes operate, and why mis- and disinformation is driven by government actors or why it is shared by individuals. It can also help reveal why people's trust in news media and their perceptions of the trustworthy new media organizations vary as much as is indicated by large-scale surveys.

Longitudinal comparative studies with global coverage are needed to assess changes in news media trust, political polarization and mis- and disinformation, using both quantitative and qualitative methods. This also applies to research on the role of AI systems and datafication in contributing to the instability of democratic institutions. To advance research that supports healthy information ecosystems and democracy, greater emphasis is needed on interdisciplinary approaches that incorporate diverse research methods and focus on technology's affordances, as well as the practices of states, companies and other key actors.

3.4 RESEARCHER ACCESS TO DATA

Research in all the fields addressed in this report is limited by problems in accessing real data (in contrast to simulated data). This limits research on decisions in the AI systems development chain, on revenue flows in the news industry and on datafication processes. Legal and ethical issues around the collection and analysis of personal and pseudo-anonymized data create barriers to data collection, and access to government and corporate data is limited in many jurisdictions, although steps are being taken to address this, for example in the European Union.⁶ A review of voluntary commitments by OpenAI, Google, Anthropic, Inflection, Meta, Midjourney and Cohere suggests slow progress in providing public application programming interfaces (APIs), deep access to data and policies for researcher access. One study concludes that AI model developers retain exclusive control 'over the majority of research access initiatives'.⁷

⁶ Forum on Information and Democracy (2024c). See, for example, Article 40, 'Data access and scrutiny', of the Digital Services Act (DSA), which sets out the conditions for 'vetted researcher' access to data of large online platforms or large online search engines, as designated under the DSA, for research that contributes to the 'detection, identification and understanding of system risks in the Union' and to the 'assessment of the adequacy, efficiency and impacts of the risk mitigation measures' (EC, 2022c). These provisions for designated platforms and the AI Act require data access, but details on data quality are controversial. See Saurwein & Spencer-Smith (2020); van Drunen & Noroozian (2024).

⁷ Harrington & Vermeulen (2024, p. 35).

There is an urgent need for safe harbors for researchers and clear data disclosure policies established by data access frameworks. A network of multinational and interdisciplinary research centers dedicated to the study of mis- and disinformation, operating in partnership with online platforms, may help to address the data access problem in countries where frameworks are not in place.

3.5 RESEARCH INDEPENDENCE

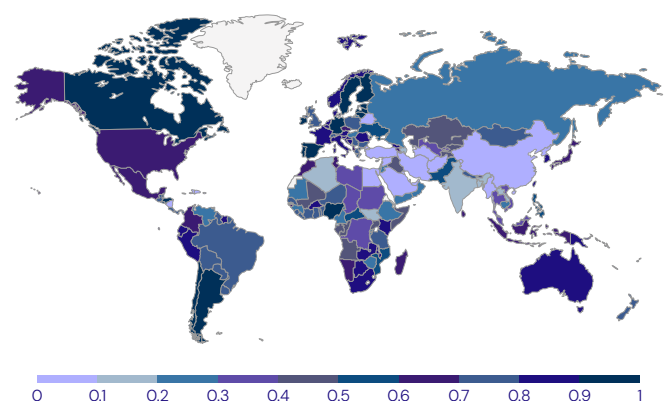
Securing the independence of researchers is vital to the validity and transparency of results. Independence from political interference and from corporate pressures to interpret research findings in ways that favor particular interests, and pressures to prioritize certain research questions, manifests in many ways. Independence is also vital for ensuring that policy makers have access to diverse perspectives and evidence. Our analysis highlights the importance of maintaining the independence of studies of news media trust (and especially studies of the financial sustainability of the press), assessments of the effectiveness of content governance methods, of audits of the performance of AI systems and of data governance frameworks and practices.

In this report we indicate in footnotes the financial support for cited research when it comes from regional/national research funding agencies, companies and a variety of civil society and industry associations when it is declared by the authors of peer-reviewed publications or reports. The intention was not to signal that cited research findings are subject to undue pressure, but that the results should be scrutinized to ensure they are not influenced by interested parties when they are relied upon as evidence.

The independence of research can be challenged when results are deemed to be politically sensitive or to question claims of companies. The politics of conducting research on information ecosystems are revealed when the causes and consequences of online mis- and disinformation are disputed in and outside research institutions.

In some countries 'scholars and students are frequently persecuted, arrested, or tortured for their academic work, research, and publications; in others, the threats to academic freedom are more subtle, often driven by market dynamics and the increase of a corporate governance model of the university'.⁸ The overt or subtle silencing of researchers often occurs when the issue is what counts as verifiable knowledge.⁹ The reality is that 'academic freedom globally is under threat', with 3.6 billion people living in countries where academic freedom is completely restricted.¹⁰ There is also evidence indicating that increased political polarization correlates positively with levels of academic freedom. Figure 9.3 shows the state of academic freedom in 2023.¹¹

Figure 9.3
State of academic freedom, 2023



Source: Kinzelbach et al. (2024, p. 1).

Note: 0–1 scale and color coding indicate low to high freedom

⁸ Furstenberg et al. (2020).

⁹ The European Parliament's Academic Freedom Monitor 2023 notes that, 'From a global perspective, the state of academic freedom in the European Union is relatively high on average compared to other regions and stable over time. Taking the European Union Member States as a reference point, there are nine countries within the European Union with a below-average level of academic freedom. These are Austria, Malta, Romania, Croatia, Bulgaria, the Netherlands, Greece, Poland and Hungary' (EP, 2024, p. 6). Until recently, in Latin America, many universities have retained autonomy, despite authoritarian governments.

¹⁰ Global levels of academic freedom increased to a peak in 2006, but by 2024 declined to their 1973 level. In 2023, only 14.1% of countries were classed as fully free, and 45.5% as completely restricted; see Kinzelbach et al. (2024, p. 1), funded in part by the Volkswagen Foundation (VolkswagenStiftung).

¹¹ Combined with the domination of academic publishing by big publishers, inequalities in knowledge production about mis- and disinformation are likely to escalate, as publishers such as Elsevier, Wiley and Taylor & Francis acquire open access repositories (Posada & Chen, 2018), and as evidence accumulates that scholars are being pressured to retract or revise their conclusions by journal editors (Teixeira da Silva, 2021).

In the *United States*, for example, researchers have been charged with using unethical research methods and results have been censored when election integrity and the role of mis- and disinformation is examined. Claims that research favors certain political parties circulate in the news media and via ostensibly independent organizations, in some cases leading to congressional investigations and undermining public trust in information that circulates in the public sphere.¹² Stanford University's Internet Observatory received pressure for undertaking rapid response tracking of electoral information.¹³ A Supreme Court ruling in August 2024 was needed to clarify whether the American government is permitted to communicate with researchers and with social media companies when they undertake research on mis- and disinformation about elections and vaccines.¹⁴ In *Brazil* researchers at the Federal University of Rio de Janeiro's NetLab were targeted by the government in an effort to decrease their autonomy in setting their own research agenda on disinformation.¹⁵

Controversy about the independent status of researchers and their institutions is not new, and it extends to research on the technical features of information infrastructures when they are deemed to raise national security concerns.¹⁶ Companies such as Google and Meta engage in charitable giving to universities in ways that may be seen as influencing research priorities, or they seed doubt in research findings.¹⁷ Retaliation against employees who become whistleblowers at companies such as OpenAI is well-documented, for example Timnit Gebru, whose research on AI ethics and facial recognition at Google was suppressed, forcing her to leave the company.¹⁸ Research evidence also suggests that the AI systems research field operates as an 'economy of virtue' where 'reputations are

traded and ethical practices are produced in line with commercial decision-making'.¹⁹ When academic access to industry AI systems is reduced, this makes it difficult for researchers to interpret industry AI models or to devise public interest alternatives.²⁰

A recurrent theme is the need to monitor the independence of researchers and their institutions and the impacts of corporate (and in some countries, government) funding.

4 Chapter Focus and Organization

Here we explain the structure of the chapters in this report. Chapter 1 provided an overview of the core themes and issues addressed, and definitions of the key concepts used. The rest of the report is structured to introduce readers to the research, focusing principally on news media and trust (Chapter 2), the development of AI systems and the implications for human rights (Chapter 3), and the generation and circulation of data within information ecosystems (Chapter 4). The research in these three chapters draws on a variety of theories and empirical evidence on the causes and consequences of changes in information ecosystems, and on the role of mis- and disinformation in changes in the conduct of debate in the public sphere.

Chapters 5 to 8 cut horizontally across the themes to address the public's and policy makers' understanding of issues and controversies and

¹² American Sunlight Project (2024), an independent organization.

¹³ In 2024 the US Department of Justice brought a case against Georgia Tech in relation to its cybersecurity lab for refusal to comply with Department of Defense (DoD) security protocols while carrying out DoD-commissioned research; see Abdalla & Abdalla (2021), supported by public money, with a third from industry; DiResta (2024); Menn & Nix (2023); Newton & Schiffer (2024).

¹⁴ Tollefson (2024). The case was initially filed by the then attorneys-general of Missouri and Louisiana, both of whom had challenged whether President Biden had won the 2020 election.

¹⁵ Medronho (2024).

¹⁶ For example, at Georgia Tech security protocols were not implemented for a period of time because they were deemed to compromise the software used in research, leading to charges brought by the Department of Justice; see Anderson (2024); Mueller (2024).

¹⁷ Graham (2024).

¹⁸ Hao (2020); Knight (2024).

¹⁹ Phan *et al.* (2022, p. 130); see also Eastwood (2024).

²⁰ Ahmed *et al.* (2023, p. 885).

research on literacy training – MIL and AI literacy (Chapter 5), legislative and regulatory measures establishing rules and norms for conduct in information ecosystems (Chapter 6), specific measures (fact-checking, industry self-regulation and co-regulation to mitigate harms of mis- and disinformation (Chapter 7), and individual and collective efforts to imagine and practice data governance consistent with fairness and justice (Chapter 8).

Here are the highlights of Chapters 2 to 8.

Chapter 2: News Media, Information Integrity and the Public Sphere.

This chapter examined what research tells us about changes in legacy and online news media, and what can be done to promote information integrity and a democratic public sphere. *What are the market structures in the news media industry and the power relations between news media organizations and digital platforms? What is the relationship between news media, a healthy public sphere and democracy? What strategies are available to the journalism profession to work towards building trust in the news media?*

The analysis covered research on the structural characteristics of news media markets and platformization, motivations to produce and consume mis- and disinformation and resilience, news media trust and distrust, the trustworthiness of legacy and online news outlets, news consumption and avoidance habits, the weaponization of information and political polarization.

The analysis highlights why market concentration and platform dominance of advertising markets contributes to the financial instability of news media organizations, how these factors affect people's trust (or mistrust) in news media content, and country differences in perceptions of the trustworthiness of news media organizations. Evidence on mis- and disinformation and political polarization is examined, demonstrating inconsistent findings about the causal effects of exposure to, and engagement with, these kinds of content.

The analysis points to the importance of strengthening the bargaining power of news organizations against platforms, differences

in findings related to factors influencing news avoidance, and ability to discriminate between accurate and false information. It calls attention to the need to extend research beyond far-right groups to government bodies, ruling political parties and others that manipulate and weaponize information during election periods. Findings are discussed, indicating that self-imposed filter bubbles can help protect marginalized groups by providing a safe space, and based on our review, that partisan online echo chambers are generally found to be smaller than typically assumed in policy debate. It points to the need for research including studies that take account of the role of legacy news media as well as online news media and political actors, longitudinal studies with global coverage to examine changes in media trust and in political polarization, and to independently monitor the news media industry's capacity to sustain trustworthy news.

Chapter 3: Artificial Intelligence, Information Ecosystems and Democracy.

This chapter examined research on the properties of AI systems (specifically machine learning (ML) algorithms) and their embeddedness in online content governance systems. *How is 'artificial intelligence' (AI) defined, and what are the relationships between AI systems development and internationally protected human rights? What are the interdependencies between AI systems development, the use of automated tools and democratic processes?* The analysis covered research on the relationships between AI systems and human rights, AI systems use and content governance (generation and moderation), and how these developments are related to changes in democracy, societal resilience and cohesion.

This chapter demonstrates how AI systems development and use are co-evolving with the safeguarding of internationally protected human rights and fundamental freedoms. It explains why states need to ensure that their obligations to respect, protect and implement these rights are responsive to specific challenges posed by new actors, instruments and power relations. The analysis highlights that no algorithm or training data set can be free of bias, and that understanding the properties of AI systems is essential if known biases

are to be mitigated. Researchers need to be specific about the algorithms or ML and LLMs that are being examined. There is substantial evidence that the use of AI systems in content governance can lead to rights violations, and modifying content governance practices ignores the multifaceted underlying causes of social discord and distrust that give rise to polarized public opinion.

The analysis concludes that no single content moderation technique will be acceptable to every online participant, which means a strong emphasis is needed on content moderation policies based on multifaceted approaches. AI tools must be used transparently and ethically, and it should not be assumed that AI systems will necessarily enhance newsroom efficiency and productivity. Measuring the scale of mis- and disinformation and its impacts is challenging without access to real data, and there is a need to provide legal frameworks for defining and removing illegal content, assuring accountability and transparency for problematic content, and rules for algorithmic personalization systems. Research needs to include an insight into how human rights law is being interpreted and applied at the country (regional) level, to assess whether commitments to protect fundamental rights are being met, to work on ways to improve data diversity, to research the conduct of (independent) algorithmic audits and, crucially, to address emerging AI divides.

Chapter 4: Big Tech Power and Governing Uses of Data. This chapter examined the relationships between the power of big tech companies and approaches to governing data extraction and processes of datafication. *What is the appropriate role of data and digital infrastructures within political communities? How are data aggregation and AI systems changing the way people build, share and receive information and knowledge? How do these big tech strategies and practices interfere with political deliberation, which is essential for the survival of participatory democracy?* The chapter provided an assessment drawing on insights into the political economy of datafication processes, which included research on digital infrastructure contestations, big tech monopolization practices and business models, and the need to work towards democratic forms of data governance.

This chapter reveals injustices associated with the interplay of data extraction and data brokering, and how digital platform business models drive data-intensive economies and a labor market that incentivizes the production of mis- and disinformation. It provides an insight into how powerful (monopolistic) actors within social, economic and political systems determine what data is produced and how it is produced, and the extent to which data is collected in ways that few understand or have control over. It analyses research indicating that much data governance legislation is permissive in fostering the amplification of mis- and disinformation and the entrenchment of global data dependencies. It emphasizes the need for research on how extractive data production has harmful consequences in people's lives, replicating and exacerbating inequalities and injustices. It also addresses data governance frameworks in countries in the Global Majority World that aim to resist the power of big tech companies; on how big tech business models make them attractive targets for mis- and disinformation campaigns; and how online labor markets incentivize the production of mis- and disinformation.

Chapter 5: Awareness of Mis- and Disinformation and the Literacy Challenge. This chapter focused on people's knowledge about the presence of mis- and disinformation in information ecosystems and literacy training initiatives aimed at enabling children and adults to identify these types of information and to protect themselves from harmful consequences. *What is known about the scale and severity of mis- and disinformation? How aware are the public and policy makers of the risks and harms of mis- and disinformation? What are the approaches to media and information literacy (MIL), and AI literacy, and what is the evidence on their effectiveness?* It provided an assessment of research in the context of the need to protect the fundamental human rights of both children and adults.

This chapter highlights challenges in measuring the severity of harms of mis- and disinformation to individuals and society in the absence of access to platform data, the tendency to neglect how

conditions in people's offline lives play a role in their experiences of online engagement, and the fact that large-scale studies are limited to a few platforms that are largely centered on the United States.

Research on public awareness of the role of AI systems in generating and circulating mis- and disinformation is shown to reach different conclusions depending on the criteria used and the context, and reveals considerable uncertainty about people's acceptance of interventions by governments or companies to tackle mis- and disinformation. It also neglects analysis of what policy makers understand about the many factors contributing to an 'information crisis'.

It highlights research demonstrating that MIL and AI literacy initiatives need to focus on more than technical skills and should include training in critical literacy; that these initiatives are not a sufficient response to mis- and disinformation; that more research on children's susceptibility to mis- and disinformation is needed to protect the rights of children; and that AI literacy training (and data or algorithmic literacy) are crucial at all stages of AI systems development and deployment. It explains why standardized MIL and AI literacy definitions and cross-country comparative conceptual frameworks and methodologies are needed, and the need for research on how critical literacy skills training can be taught effectively to children and adults.

Chapter 6: Governing Information Ecosystems: Legislation and Regulation. This chapter provided an account of selected legislative and regulatory tools available to governments to mitigate the harms of mis- and disinformation, and to govern the way mainly big tech companies operate. *What types of governance approaches are available? What approaches to information ecosystem governance are being promoted at the global level? What are some of the legislative, regulatory and judicial approaches to governing information ecosystems?* This chapter emphasized normative goals and rules embodied in governance approaches, providing an insight into tensions between these goals and their implementation in view of the interests of different actors. The analysis focuses on principles and

guidelines reflected in legislation and regulations with respect to network infrastructure, privacy and data protection, digital platforms, AI systems and news media.

This chapter highlights variations in governance measures around the world, especially on the penalization or criminalization of those who produce and circulate mis- and disinformation. It draws attention to research demonstrating why attention to network neutrality policies and 'zero-rating' regulations is crucial, and why human rights principles should guide normative expectations for the role of the news media, even if deviations occur in practice, highlighting that regulation applied to legacy and online news media can result in censorship or leverage over news media organizations. It explains why privacy and data protection legislation is not a panacea for all data economy issues, and why homogeneous approaches to governing AI systems and tackling mis- and disinformation are not likely to be viable. There is a need for research to monitor voluntary and legal governance measures; to track corporate lobbying; to assess whether measures are helping people navigate information ecosystems in ways that enhance resilience to mis- and disinformation; and to assess whether governance is aligned with both individual and collective interests and with experience in the Global Majority World.

Chapter 7: Combating Mis- and Disinformation in Practice. This chapter looked in detail at specific governance measures to combat mis- and disinformation undertaken by civil society organizations and introduced by governments. *What content governance efforts are being made to combat mis- and disinformation? What are the challenges in achieving effective information ecosystems governance? In what ways are human rights protections jeopardized by governance aimed at curtailing online mis- and disinformation? What is known about the public's appetite for interventions to moderate online mis- and disinformation?* The analysis emphasizes the need to differentiate between the stated aims of governance and its consequences when practice falls short of normative expectations. It focuses on fact-checking, industry self-regulation,

co-regulatory approaches and the public's view of how mis- and disinformation issues should be addressed.

This chapter calls attention to research indicating that human rights protections are jeopardized by some measures to combat mis- and disinformation, and the need to achieve greater clarity about intervention points where governance can have positive impacts, illustrating why a single approach is neither feasible nor desirable. There is a need for a shared understanding of the appropriate balance between the imperatives of economic growth, innovation and human rights protections as well as guarding against regulatory overreach, particularly by authoritarian governments. It highlights an overemphasis on technical tools to support mis- and disinformation countermeasures rather than on diverse contexts, emphasizing that practices such as fact-checking are not static processes, so their effectiveness is likely to vary over time. It draws attention to differences between countries in the way they seek to protect press freedom and to counter online mis- and disinformation. Future research is needed to test countermeasures with real-world data beyond Global North countries. The chapter discusses the benefits of using mixed methods to reveal a range of experiences, and the need to monitor digital platform practices that result in the suppression of voices that are critical of state authorities.

Chapter 8: Towards Data Justice in Information Ecosystems. This chapter examined research explaining how the monopolistic power of big tech companies creates biases and harmful discrimination and exclusions, infringing on people's human rights in a data economy that thrives on data extraction and monetization. *Why do corporate incentives, strategies and practices involved in designing, developing, selling and controlling data lead to epistemic injustice? What strategies and tactics are individuals and communities developing to resist the extractive features of the data economy?* This chapter emphasizes individual and collective dependencies and inequities resulting from datafication, and how datafication practices might be reimagined to empower individuals and communities in ways that contribute to data

justice. It focuses on the consequences of biased AI systems for human rights guarantees and democratic decision making, and individual and group (local, municipal and national) resistance strategies to current practices.

This chapter assesses research demonstrating that commercial datafication supported by AI systems disadvantages and discriminates among people in the data economy by sustaining comprehensive surveillance to enable computerized data production and services. It highlights the epistemic injustices (the privileging of information and knowledge that are neither representative nor inclusive) and the individual and collective dependencies and inequities resulting from datafication, including the consequences of biased data on which AI systems are trained. It reviews research on initiatives taken by individuals and groups to think critically about how to govern massive amounts of digitized data, and highlights strong pressures from civil society to treat data governance as a lever for restructuring data markets, to protect against infringements of human rights and to tackle concentrations of power and wealth that jeopardize democracy. Future research must work on decolonizing knowledge about and experiences of the data economy, monitor discriminatory outcomes of datafication and examine how dependencies on big tech companies are created. It emphasizes the need for greater insight into strategies to advance public interest alternative news media, Indigenous community and municipality initiatives, and develop both community-controlled technologies and decentralized data governance frameworks.

5 Limitations of the Report

This critical analysis of state-of-the-art research on important components of information ecosystems is limited in several ways, which are set out in detail in Chapter 1. Briefly, they include:

- A focus on material inequalities in people's lives only to the extent that broad socio-economic conditions are mentioned, since the analysis is focused on the themes and questions that structured the analysis.
- A principal focus on the upper service applications layer of information ecosystems, although several issues on the infrastructure layer that affect the health of information ecosystems and the quality of debate in the public sphere are discussed.
- An imbalance in Global North and Global Majority World research sources favoring the Global North, notwithstanding our efforts to reach out to be more inclusive.
- The analysis does not aim to cover the extensive research on 'digital divides', although we acknowledge huge variations in the availability of meaningful internet connectivity and access as well the presence of restrictions on access to information.
- This analysis does not cover research on cybersecurity, securitization, geopolitics and 'digital sovereignty' or the economic geography of digital labor markets or the (micro)economic analysis of digital markets.
- The focus tends to be on country-level experience and institutions, with no attempt to include micro-level or sectoral experience, technology 'use cases' or 'case studies'.
- Analysis in this report is inevitably limited by the fact that all research is guided by research questions selected for investigation by research communities, the funding available to do research and the researchers' access to data.

6 A Final Word on what should be done

Our critical analysis of state-of-art-research amply demonstrates a privileging of knowledge about information ecosystems produced in and about the Global North. It also confirms that the affordances of digital systems (including AI systems) are complicit in failures to protect human rights in the Global North and Global Majority World. There is controversy in the research literature about the principal reasons for this – for example, whether the norms and practices of monopolistic companies and states or individual behaviors and attitudes are the predominant explanations for the spread of viral mis- and disinformation. There is an absence of consensus in research evidence about how best to tackle harms associated with mis- and disinformation, and the wider issues around the fragility of democracy. Controversy partly arises from differences in the way problems are identified, conceptualized and studied. Controversy is also attributable to distinctive cultural, social, political and economic conditions in countries around the world.

6.1 ROLE OF CIVIL SOCIETY ORGANIZATIONS

Despite these controversies, this report has identified where future research can help to address them. This report is mainly based on academic research, but it also benefits from research undertaken or commissioned by civil society organizations and other non-university independent non-profits (we cite 118 of these – 47% Global North, 27% Global Majority, 26% Global coverage).

Civil society organizations play a vital role in 'speaking truth to power' and responding to the exploitative data practices of big tech companies and, in some instances, governments. Acknowledging this role involves:

- Recognizing that civil society groups are working with academic researchers to call attention to these practices, and thinking critically about

how to devise just data governance practices and how to build alternative data governance frameworks.

- Encouraging civil society groups and academics who are working on local, community and municipal data governance frameworks and on proposals to introduce decentralized data governance at the national level.

Our report also highlights areas where actions could be taken by governments or the private sector – actions aimed at ensuring that strategies and practices are consistent with international human rights commitments and with strengthening a democratic public sphere. This report was not designed to generate specific recommendations for policy makers or companies, but guidance is provided in this section.

6.2 GUIDANCE FOR POLICY MAKERS

It is important to acknowledge the limits of policy action in the face of corporate power, divided publics, and current political institutional norms and practices, but policy makers can take steps to promote healthier information ecosystems by learning from research evidence. For governments, some actions require new or different governance measures. Others are about how policy makers think about information ecosystem problems, information integrity, the role of new technologies and the problems created by mis- and disinformation.

Tackling Power Asymmetries

- Unhealthy information ecosystems are clearly facilitated by big tech monopolistic business strategies that encourage commercial data monetization. A comprehensive systemic approach is needed if policy makers are to tackle what is widely seen as an ‘information crisis’ that threatens democratic stability.
- Policy makers should deploy the full range of governance approaches available to them including co-regulatory approaches and competition/anti-trust measures to restrain the big tech industry’s use of business models

that lead to the amplification of mis- and disinformation and harms to children and adults.

- Evidence indicates that the challenges of governing foreign-owned big tech companies can be addressed by encouraging coalitions of country or regional stakeholders that work to counter the power of these companies.
- Policy must address structural inequalities in digital services markets and political alignments that foster mis- and disinformation which destabilize democracy, especially those that prevent news media independence and stand in the way of treating news media as a ‘public good’.
- Steps must also be taken to reimagine and foster alternative datafication models aligned with data justice principles. This means supporting initiatives to build alternative data governance frameworks including local, community and municipal and decentralized national data frameworks and incentivizing the work of civil society organizations that monitor big tech data harmful practices and work to reimagine alternatives.
- It is essential that policy makers preserve and promote the capacities of diverse communities to question dependencies on the products and services provided by big tech companies outside formal policy-making spaces as well as through participation in formal consultative processes.

Independently monitoring human rights infringements

- Investment in monitoring human rights infringements associated with information ecosystems is essential. Evidence indicating that the interests of big tech companies are being favored in policy decisions (even when legislation is in place) due to weak enforcement must lead to steps to put more effective governance in place.
- Policy makers must recognize that measures to combat mis- and disinformation risk suppressing voices that are critical of state

authorities and take steps to ensure that these risks are mitigated.

- Research demonstrates an urgent need for policy measures to secure the safety of journalists and to limit political pressure and other constraints on media freedom especially when these pressures give rise to mis- and disinformation with severe offline impacts.
- Studies highlight the features of datafication strategies that place disproportionate burdens on marginalized populations and these need to be addressed urgently.

Measures to combat mis- and disinformation

- Policy measures are needed to hold big tech companies accountable for the services and AI tools they release to the market. This means monitoring the growing use of personalization systems and AI tools, including GenAI tools.
- Evidence demonstrates that no single content moderation technique will be acceptable to every online participant. This means recognizing that multiple approaches are needed to combat mis- and disinformation, rather than relying disproportionately on AI tools.
- Policy measures are needed to address the financial instability of the news industry in many countries, to promote independent news media and to counter the dependence of news media organizations on digital platforms. This means addressing big tech company resistance to making ad tech revenues transparent, devising ways to ensure independent public service media and smaller local news outlets are financially viable to protect media freedoms and a plural and diverse public sphere.
- Evidence shows that where trust is declining in news media (and public institutions) this cannot be addressed solely by promoting the use of AI systems and other technical measures or by promoting fact-checking. These measures need to be complemented by policies targeting the incentives created by big tech business models.

- It is essential to encourage investigations of the actors and institutions that generate mis- and disinformation and their motivations. These tend to be neglected in policy that favours efforts to mitigate individual harms and a more balanced approach could help to counter the production of mis- and disinformation and its circulation.

Strengthening Transparency and Accountability

- Policy makers must ensure that big tech companies provide fully transparent reports, for example, on content moderation processes (including personalization algorithms), known algorithmic biases, third-party data sharing agreements, and data breaches. This is key to understanding whether these companies are being held to account. Enforcing AI system transparency by ensuring regular independent audits is crucial.
- Research demonstrates that accountability and transparency measures applied to big tech companies can be weakened when barriers exist between state and regulatory institutions charged with implementing them. Policy makers should take steps to improve policy coordination especially for policy aimed at countering mis- and disinformation.
- Evidence indicates that policy coordination is especially important to enforce measures aimed at governing political campaigning and political spending which fosters mis- and disinformation.

Media and Information Literacy (MIL) and AI Literacy

- Media and information literacy (MIL) and AI literacy training for adults and children is a promising means of granting people greater control over their information environment. Policy makers can foster measures to encourage additions to education curricula or encourage private sector and civil society coalitions to provide training and evaluate outcomes over time.
- Recognizing that these training programs are essential to enable children and adults to

interpret and critically the information they are exposed to online is a crucial step.

- It is also essential to evaluate these initiatives over the long term and to adequately resource them.
- MIL or AI literacy training should not be treated as the main solution to unhealthy information ecosystems and declining trust in news media. A systemic approach is needed to address the factors contributing to unhealthy ecosystems.

Influencing Research Priorities

- Frameworks need to be put in place to provide useable data for research purposes. This means implementing frameworks for researcher access to data, ensuring that these are respected, and monitoring concerns of the research community about their adequacy.
- Policy can encourage global cross-disciplinary, collaborative and comparative research through multinational research centers, including the Global Majority World, to examine the incidence and multiple causes of mis- and disinformation. This is likely to require a multinational and interdisciplinary network of research centers operating in partnership with those big tech companies that are willing to acknowledge and examine how data access policies and practices influence what research is undertaken and its results.
- Policy can incentivize multidisciplinary research that joins up work on mis- and disinformation, political processes and market structures with research on cybersecurity and geopolitical tensions. This is essential to capture the interdependency of the components of information ecosystems and their outcomes for individuals and society.
- Policy can encourage research that moves beyond the laboratory to test AI system based methods of detecting and combating mis- and disinformation employing a variety of quantitative and qualitative research methods.

- Policy can help to sustain a public infrastructure for independent research and ensure that research findings are not suppressed for political reasons.
- Policy makers should encourage the inclusion of more diverse types of research in the evidence base that is used to inform policy. This means including research on the socio-economic and political conditions in society that give rise to mis- and disinformation and on the structural and power relations in the big tech industry as well as the impacts on individual attitudes and behaviors.

6.3 GUIDANCE FOR BIG TECH COMPANIES

Big tech companies make public commitments to promote safe and democratic online spaces. These commitments require that companies:

- Change their business strategies and uses of technologies in view of the longer-term negative reputational effects of their data monetization models which harm individuals and groups and are linked to democratic instability.
- Introduce strategies and practices that are fully aligned with international human rights commitments including the rights of the child.
- Provide fully transparent reports voluntarily in countries where legislation is not in place, or in response to legislative requirements, for example, on content moderation processes (including personalization algorithms), third-party data sharing agreements, data breaches, measures to address known algorithmic biases and provide useable data for research purposes.
- Invest in inclusive mechanisms for consulting with individual users and collective organizations about their experiences of mis- and disinformation and their participation in the digitized public sphere and attending to responses that they find acceptable.
- Adequately resource content moderation processes and ensure that the conditions for

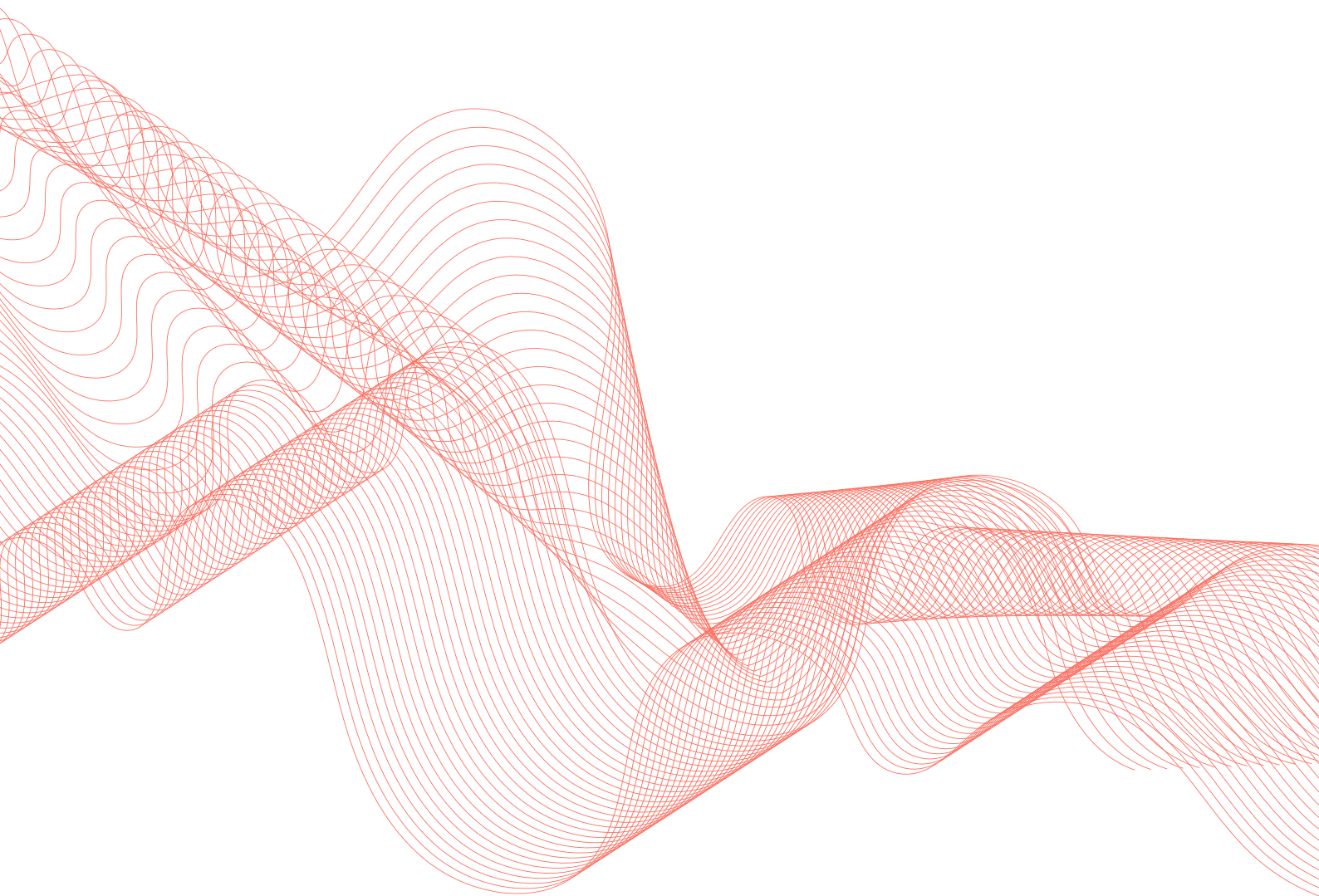
workers meet acceptable standards of pay,
health insurance and care for mental health.

- Increase transparency by report lobbying expenditure and reporting on which topics – regulatory procedures and court cases – are the target of lobbying activities.

It is crucial to encourage initiatives from all stakeholders – corporate, government, civil society organizations, philanthropic organizations and academics – if the United Nations' *Global Digital Compact* goal of promoting 'diverse and resilient information ecosystems' is to be met. These initiatives will have a greater chance of success if they are based on an understanding of information ecosystems that recognizes their complexity, of how they are developing and experienced differently depending on the context, and on whether government and corporate practices are successful in upholding international human rights commitments.

APPENDIX

METHODOLOGY



1 Introduction

The International Observatory on Information and Democracy, a core project of the Forum on Information and Democracy, aims to provide a common and shared understanding of information ecosystems and their impact on democracy.

The Observatory aggregates and synthesizes existing research and available data through a robust, inclusive critical review process. Currently in the form of biennial reports, it provides civil society leaders, researchers, academics and, importantly, policy makers, with a periodic global assessment of the information and communication space and its impact on democracy.

Its work aims to inform the international community's efforts to foster the adoption of effective and proportionate regulatory and non-regulatory measures for the protection of human rights – including the right to reliable information – and democracy in the digital space.

The sections of this appendix outline important aspects of the Observatory's work that led to the production of this report. Section 2 deals with the governance structure and the configuration of the Observatory's work from 2022 to 2025. Section 3 explains the data collection phase and the critical state-of-the-art review methodology, including the criteria for selection and analysis of the literature. Section 4 provides a breakdown of the citations used in the report – what literature is analyzed – and provides an insight into the limitations of the work. Finally, Section 5 details the reasoning for the terminology used in the report, specifically the concepts of 'information ecosystem' and 'public sphere'.

2 The Observatory's Work and Governance

2.1. PREFIGURATION PHASE AND GOVERNANCE STRUCTURE (2022–23)

The **prefiguration group** of the International Observatory on Information and Democracy, composed of recognized personalities from the world of research and international governance (listed below), was tasked to specify the objectives, methodology and means of the Observatory in a report, *Observatory on Information and Democracy – Feasibility study*. The report was presented to the States of the International Partnership for Information and Democracy during the 2022 Summit on Information and Democracy.

Co-Chairs

- Ángel Gurría, former Secretary-General of the OECD
- Shoshana Zuboff, Professor Emeritus, Harvard Business School, author of *The Age of Surveillance Capitalism*

Members

- Virgílio Almeida, Professor Emeritus, Department of Computer Science, Federal University of Minas Gerais (Universidade Federal de Minas Gerais)
- Jim Balsillie, Founder of the Center for International Governance Innovation (CIGI)
- Jean-Marie Guéhenno, diplomat, former United Nations Under-Secretary-General

- Miguel Poiaras Maduro, Chair of the Executive Board, European Digital Media Observatory, European University Institute
- Maria Ressa, Chief Executive Officer of Rappler, Nobel Peace Prize laureate 2021
- Burhan Sönmez, President, PEN International

In May 2023, more than 100 researchers expressed support for the creation of the Observatory and called for all relevant stakeholders to cooperate with it. On the International Day of Democracy, the prefiguration group nominated 19 prominent leaders in policy, research and academia to join the **Steering Committee** for the Observatory's inaugural research cycle. This dynamic group of scholars and thought leaders (listed below) represents diverse geographies and disciplines, encompassing fields as wide-ranging as political science, ethics, journalism, engineering, anthropology, economics and data science. The job of the Steering Committee was to take part in shaping the content and presentation of findings in the report through active feedback on the research method and report drafts.

- Luca Belli, Professor of Digital Governance and Regulation, Fundação Getulio Vargas (FGV) Law School, Rio de Janeiro (Brazil)
- Gustavo Cardoso, Media Sociologist, University Institute of Lisbon (ISCTE-IUL, Instituto Universitário de Lisboa), Director of OberCom (Observatório dos Meios de Comunicação Social) (Media Observatory) (Portugal)
- Niva Elkin-Koren, Professor at Tel-Aviv University Faculty of Law, Faculty Director of the Chief Justice Meir Shamgar Center for Digital Law and Innovation (Israel)
- Helani Galpaya, Chief Executive Officer of LIRNEasia (Sri Lanka)
- Timothy Garton Ash, Professor of European Studies, University of Oxford, Isaiah Berlin Professorial Fellow at St Antony's College, Oxford, (United Kingdom); Senior Fellow at the Hoover Institution, Stanford University (United States)
- Frances Haugen, Co-Founder of Beyond the Screen, algorithmic product expert, advocate for accountability and transparency in social media (United States)
- Jeanette Hofmann, Professor at Freie Universität Berlin, Research and co-founding director of the Alexander von Humboldt Institute for Internet and Society (HIIG) (Germany)
- Jhalak Kakkar, Executive Director, Centre for Communication Governance, National Law University Delhi; Visiting Professor, National Law University Delhi (India)
- Ansgar Koene, EY Global AI Ethics and Regulatory Leader (Belgium)
- Admire Mare, Associate Professor and Head of Department, Department of Communication and Media Studies, University of Johannesburg, South Africa (Zimbabwe)
- Nnenna Nwakanma, digital policy, advocacy and cooperation strategist (Ivory Coast)
- Pier Luigi Parcu, Director of the Centre for Media Pluralism and Media Freedom, European University Institute, Director of the Centre for a Digital Society, Robert Schuman Centre (Italy)
- Courtney Radsch, Director of the Center for Journalism and Liberty, Open Markets Institute, Fellow at UCLA Institute for Technology, Law and Policy, Fellow at the Center for Democracy and Technology (United States)
- Jeremy Rifkin, economic and social theorist, bestselling author of 23 books, President of the TIR Consulting Group LLC (United States)
- Ghassan Salamé, former Minister of Culture of Lebanon, Emeritus Professor of International Relations at Sciences Po Paris (Lebanon and France)
- Sonja Solomun, Deputy Director of the Centre for Media, Technology, and Democracy, McGill University (Canada)
- Nicol Turner Lee, Senior Fellow in Governance Studies, Director of the Center for Technology Innovation, Brookings Institution, Washington, DC (United States)

- Stefaan Verhulst, Research Professor, Center for Urban Science and Progress, Tandon School of Engineering of New York University, Co-Founder of The Governance Laboratory (GovLab) (NYC), Co-Founder of The Data Tank (Brussels)
- Natalia Zuazo, author of *Guerras de internet* and *Los dueños de internet*, Director of SALTO, UNESCO consultant (Argentina)

In parallel, the Observatory set up its **Stakeholder Advisory Group** (SAG) for consultative purposes. This was organized into three sub-groups representing stakeholders working in the fields of information, communication and democracy:

- States and regulatory bodies (gathering all 52 States of the Partnership for Information and Democracy and regulatory representatives)
- The world of research and advocacy (an extensive network of NGOs and academic representatives)
- Private tech companies (spanning a dozen private tech company representatives, including some very large online platforms, or VLOPs).

The SAG was consulted at key milestones of the research cycle to ensure diverse perspectives. The group made recommendations and shared best practice with the Steering Committee regarding the production of the Observatory's report.

2.2. INAUGURAL RESEARCH CYCLE (2023–25)

After the prefiguration phase that set out the Observatory's objectives, means and methodology, and a thorough consultation process with all relevant stakeholders in the field, the first work cycle was officially launched in October 2023 at the Internet Governance Forum in Kyoto, Japan. During the event, an interactive panel discussion featuring members of the Steering Committee revealed the priority **themes** of the inaugural report: **news media**, **AI systems** and **data governance**. These were to be dealt with in the context of the challenges of mis- and disinformation for the integrity of public discourse, the fairness of political processes, media freedom and the resilience of public institutions.

The Observatory issued an open call to the SAG and other key contacts to recruit members for three **Research Assessment Panels** (RAPs) that were tasked to aggregate and synthesize state-of-the-art research relating to the three themes (see Sections 3 to 5 for the detailed methodology). The permanent Secretariat of the Observatory selected the panel members based on experience and expertise in the relevant fields, regional expertise and publication records. Special attention was given to ensuring a gender balance and geographical diversity during the selection process. Initially 60 experts and researchers were selected to join the RAPs, with many contributing throughout (see the Acknowledgments).

In addition, a **scientific director**, a lead rapporteur, and two rapporteurs for each RAP were appointed to draft the report. **Robin Mansell**, Professor Emerita, Department of Media and Communications, London School of Economics and Political Science (LSE), was appointed as the Observatory's Scientific Director, and Professor **Rob Procter**, University of Warwick and The Alan Turing Institute, was named as the Lead Rapporteur.

3 Research Themes and Methods

3.1. OVERVIEW AND RESEARCH GOAL

Rather than undertaking original research, the Observatory's mandate is to conduct a critical synthesis of state-of-the-art research. As such, the report synthesizes existing evidence based on a critical review of the literature informed by expert consultations. The aggregation and synthesis of research and available data proceeded under the supervision of the Observatory's Steering Committee chaired by Courtney Radsch, Director of the Center for Journalism and Liberty, Open Markets Institute, and 18 other renowned international experts in the field.

The report summarizes existing research and significant gaps in the evidence base, identifying priorities for future research. In this effort, the Observatory especially looked to source input from the Global Majority World – albeit acknowledging and reflecting on the gross imbalance in resources available to conduct and publish research between the Global North and the Global Majority World. The report presents the key findings and actionable insights, highlighting the main differences in information ecosystems around the world and how people engage with them, based on research drawing on a wide variety of theories and research methods.

3.2. A THEME-CENTERED APPROACH

The work of the Observatory was conducted around three guiding themes: **media, politics and trust; artificial intelligence, information ecosystems and democracy; and data governance and democracy**. Theme rapporteurs and RAP contributors examined evidence relevant to the themes with a cross-cutting focus on mis- and disinformation and a particular interest in their intersection with democratic attributes. In doing so, they focused on key aspects such as the integrity of public discourse, the fairness of political processes, media freedom and the resilience of public institutions.

Each theme was addressed using a set of guiding questions and objectives set by the Steering Committee with two main goals: (i) mapping research and policy, synthesizing the strengths and weaknesses of the evidence; and (ii) recommending new lines of research and potential governance responses (formal policy/regulatory, multilateral, multistakeholder, etc.). This research exercise was not intended to provide specific recommendations for policy action.

A brief overview of each theme follows.

Media, politics and trust

This theme addressed the pressures, adaptations and impacts on reporting local, national, regional and international news, business models for the remuneration of journalism, the political implications of online mis- and disinformation for democracy in the longer term, and the impacts of mis- and disinformation on trust in science, experts and the media in a 'post-truth' era.

Objectives:

- To understand the changing role of legacy and newer news media in democratic processes, including journalist reporting and trust in media sources.
- To assess the causes and consequences of changes in news media practice for democracy and the reasons for declining trust.

Artificial intelligence, information ecosystems and democracy

This theme addressed the use of AI systems in content moderation and personalization systems in information ecosystems, their use in legacy and news media production, and the implications for the protection of fundamental rights and for the fabric of democratic societies, including values, structures, and processes resulting from changes in global communication flows and information quality. It also addressed the need for people to acquire skills to help them manage their consumption of content encountered within information ecosystems, and to participate in debates about AI applications that affect their working and everyday lives.

Objectives:

- To understand the role of AI systems development in information ecosystems.
- To assess the causes and consequences of AI systems for democracy, including the broader implications of the role of AI systems in information ecosystems for democratic processes, societal resilience and cohesion, including democratic values.

Data governance and democracy

This theme focused on artificial intelligence as an anti-democratic economic phenomenon and anti-democratic epistemic form. It critically examined the discourses and strategies deployed by big tech companies to promote their 'AI' solutions in ways that mystify what AI is and can do, creating dependence through the exercise of corporate power and curtailing opportunities for uses of AI consistent with healthy, inclusive, fair and just democracies. It also focused on how data governance approaches developed in and by actors in the Global North are exported and resisted by actors in the Global Majority World.

Objectives:

- To understand the role of data governance and its implications for fundamental human rights and democracy.
- To assess the impact of approaches to data governance, including the broader implications for democratic processes, practices and inclusive participation.

Cross-cutting theme: mis- and disinformation

This theme aimed to elicit an insight into the problems for democracy created by the production, circulation and consumption of information that qualifies under definitions of 'misinformation' and 'disinformation' by undertaking a critical analysis across the three themes. It recognized that the deployment of AI systems is reshaping the public sphere and the processes and practices of democracy, with implications for inclusive and equitable participation in society in the Global North and Global Majority World.

Objectives:

- To understand the role of mis- and disinformation in information ecosystems and its implications for democracy.
- To assess the causes and consequences of mis- and disinformation for democracy, including the broader implications for democratic processes, practices and participation.

3.3. DATA COLLECTION AND ANALYSIS

The research assessment in this report can be characterized as a critical review of state-of-the-art research. The strength of our review is our effort to critically evaluate the quality of cited works, and at the same time, our examination of relatively current research and commentary. Undertaking a formal systematic review following standards such as the Cochrane methodology was outside the scope of the project. Thus, we did not undertake a quantitative meta-analysis of the research collected for this report.¹ Here, we provide an overview of the key steps in our review.

¹ Grant & Booth (2009), for systematic literature reviews, and Higgins et al. (2024), for guidance on a formal Cochrane protocol review.

First, the Scientific Director – together with the Lead Rapporteur and the rapporteurs – compiled and discussed keywords,² which they used to search bibliographic databases and a variety of web-based sources, starting the search in October 2023 with additions made until October 2024.³ In addition, the Scientific Director searched titles and abstracts of papers in issues of 52 journals published between 2022 and December 2023, including journals in the social sciences and computer science.⁴ Additional Google Scholar, Web of Science and Scopus searches were conducted when the Steering Committee suggested sub-topics that should be explored – keywords in these cases were specific to the topic, for example zero rating, news deserts and digital divide. Earlier publications (before 2020) were included when it was appropriate to provide background or historical context.

Second, individuals or groups of two or three RAP contributors and the rapporteurs for each theme conducted additional searches guided by the objectives (see Section 3.2) and the research question (a selection of these is indicated at the start of Chapters 2–8 and in Chapter 9). The guidance issued to RAPs for the selection of sources is indicated below.

Third, two global calls for contributions were launched to gather further sources. The focus was to gather material especially concerning the Global Majority World to enhance the diversity of our citations database.⁵ The Observatory's team and rapporteurs also proactively reached out to experts based in the Global Majority World to identify further relevant sources, with the rapporteurs conducting selected interviews.

Selection criteria

Relevance: Relevance of sources to the themes.

Publication date: Publication date emphasized during the search process was from 2022 to October 2024 to capture recent work but earlier sources were included as appropriate.

Country focus: Global Majority World, Global North, with countries/regions specified as needed, and Global, work with global reach.

Document type: PhD theses/dissertations, academic peer-reviewed articles, academic books and book chapters (some books by non-academic authors active in industry), research reports, policy reports, conference papers and proceedings, pre-prints, and other (magazine and newspaper articles, blogs).

² Among the keywords searched were: AI, content generation, content governance, content moderation, mediated misinformation, AI literacy, algorithmic transparency, fairness, information behavior, algorithmic bias, democratic institution, machine learning, news personalization, recommender systems, public trust, public sphere, electoral process, algorithmic accountability, AI ethics, deepfake, news media, traditional media, legacy media, digital media, alternative media, public service media, commercial media, digital platform, social media platform, community media, media ecosystem, trust in news, media consumption habits, news avoidance, media diet, journalists, local, regional, national news, monopoly, competition, data governance, epistemic, human rights. These search terms were combined with information ecosystem, information integrity, misinformation, disinformation, malinformation, hate speech, in varying combinations.

³ Among the online repositories searched were: AAAI/ACM Conference on AI, Ethics, and Society, AAAI Conference on Artificial Intelligence, ACM Conference on Fairness, Accountability, and Transparency, ACM Conference on Human Factors in Computing Systems, ACM SIGIR International Conference on the Theory of Information Retrieval, Annual Meeting of the Association for Computational Linguistics, European Data Journalism Conference, European Conference on Data and Computational Journalism, IEEE Conference on Artificial Intelligence, International Conference on AI-generated Content, ACM International Conference on Multimedia, International Conference on System Sciences, ArXiv and SSRN Scholarly Papers.

⁴ A search on the indexes of the following journal titles and abstracts was performed using information ecosystem, misinformation and disinformation, information integrity, malinformation, hate speech, democracy, algorithm, regulation, and governance, in varying combinations, but excluding Covid-19-related papers: *ACM Computing Surveys*, *The African Journal of Information and Communication*, *Artificial Intelligence*, *Artificial Intelligence and Law*, *Artificial Intelligence Review*, *Asian Journal of Communication*, *Asian Journal of Information Technology*, *Berkeley Technology Law Journal*, *Big Data & Society*, *Big Data Mining and Analytics*, *Chinese Journal of Communication*, *Computer Law & Security Review*, *Digital Communication and Networks*, *Digital Journalism*, *Electronic Journal of Information Systems in Developing Countries*, *Energy and AI*, *European Journal of Communication*, *Foundations and Trends® in Machine Learning*, *Frontiers in Big Data*, *Harvard Journal of Law & Technology*, *Human-Computer Interaction*, *Information, Communication & Society*, *Information Development*, *Information Systems Journal*, *Innovation and Development*, *International Data Privacy Law*, *International Journal of Communication*, *International Journal of Information Communication Technologies and Human Development*, *International Journal of Information Management*, *The International Journal of Press/Politics*, *International Review of Law, Computers & Technology*, *Internet Policy Review*, *Journal of Computer Mediated Communication*, *Journal of Digital Technologies and Law*, *Journal of Interactive Advertising*, *Journal of Law & Innovation*, *Journal of the ACM*, *Journalism & Mass Communication Quarterly*, *Journalism Studies*, *Knowledge-Based Systems*, *Law, Innovation and Technology*, *Media, Culture & Society*, *New Media & Society*, *Political Communication*, *Public Opinion Quarterly*, *Social Media + Society*, *Social Science Computer Review*, *South African Journal of Information Management*, *Stanford Technology Law Review*, *Telematics and Informatics*, *The Information Society*.

⁵ It is likely that many of the academic sources in our database were written by Global Majority World scholars who were either trained or work and live in the Global North, but this was not checked systematically.

Language: English (a small selection of French, German, Italian, Portuguese and Spanish).

Type of studies: Studies that use: (i) well-established (qualitative or quantitative) social science methodology and rely on original research or secondary analysis to make an empirical or theoretical contribution to the themes; and (ii) policy reports or other types of contributions published by reputable organizations that may not adopt a clear methodology or rely on primary data, but that identify unique trends relevant to the themes.

Contributions spanning different fields of expertise and regions were assembled using Zotero's open-source bibliographic software, with a total of **3,095 sources** collected by October 2024. Sources were downloaded to Zotero and/or linked by URL. All entries in the full database were reviewed by at least one relevant rapporteur or by the Scientific Director to decide on inclusion in the report. All sources selected for inclusion were reviewed throughout the drafting process by at least two members of the drafting team, usually one rapporteur and the Scientific Director. Cited sources not available on open access were sourced through university library accounts, via drafting team book collections and, in a few instances, purchased.

Rapporteurs reviewed each entry for its type (peer-reviewed article, book, book chapter, report, other material); whether it was public or private (confidential); its origin (authorship, country or public/private organizations); and where indicated, its financial support (this was noted for national/regional research funding bodies, corporate and direct government funding, and public and private organization support; university financial support was not recorded). Most sources were in English, but a few entries were in French, German, Italian, Portuguese and Spanish, with some text translated using online support and own translations.⁶

All citations were checked for quality (see Table A1) in order to (i) maximize the use of independent research sources in the review; and (ii) note sources where funding was declared in footnotes. Importantly, sources characterized as 'low' quality were not automatically excluded, but rather assessed in the light of differences in the availability of independent funding in different countries.⁷ The RAP members then synthesized the source evidence for the rapporteurs, sometimes drafting directly text for potential inclusion in the report. Sources were assessed for the novelty of insights and their robustness in view of their provenance. (See Section 4 for details of the final collection of cited sources.)

The report team and the Observatory's office team met weekly to review progress. Two drafts of the report were presented to the Steering Committee and to the SAG for their feedback, which was then integrated into a final draft that was submitted for approval to the Steering Committee. These drafts were also reviewed by RAP members.

Despite gaps in evidence and uncertainties – some due to real-world differences in the experience of information ecosystems, others to diverse theories and methods – this critical review of state-of-the-art research provides an insight into evolving information ecosystems standards, conventions, best practice and changes in the public sphere.

⁶ Translations of portions of the texts in German, Portuguese and Spanish were performed by Google Translate; French language reading proficiency was excellent or good.

⁷ In certain countries, funding for research on information systems is often granted mainly by the government or a private foundation in the absence of other sources. In these cases, the quality of the research was assessed carefully in light of the country context by the rapporteurs and RAP members.

Table A1

Definitions of source quality

Quality	Definition	Funding
High	Paper published in high-quality, peer-reviewed journal or conference.	Independent.
	Book/report published by internationally recognized publisher or organization.	
	Book/report written by internationally recognized author(s).	
	Book chapter written by internationally recognized author(s).	
Medium	Paper published in peer-reviewed journal or conference.	Independent or not declared.
	Book/report published by less well-known publisher or organization.	Independent.
	Book chapter written by less well-known author(s).	
	Report.	Sponsored (government organization; intergovernmental organization; civil society organization).
Low	Report.	Sponsored (private company; private sector association).
	Report.	Not declared.
	Blog, newspaper or magazine article.	-

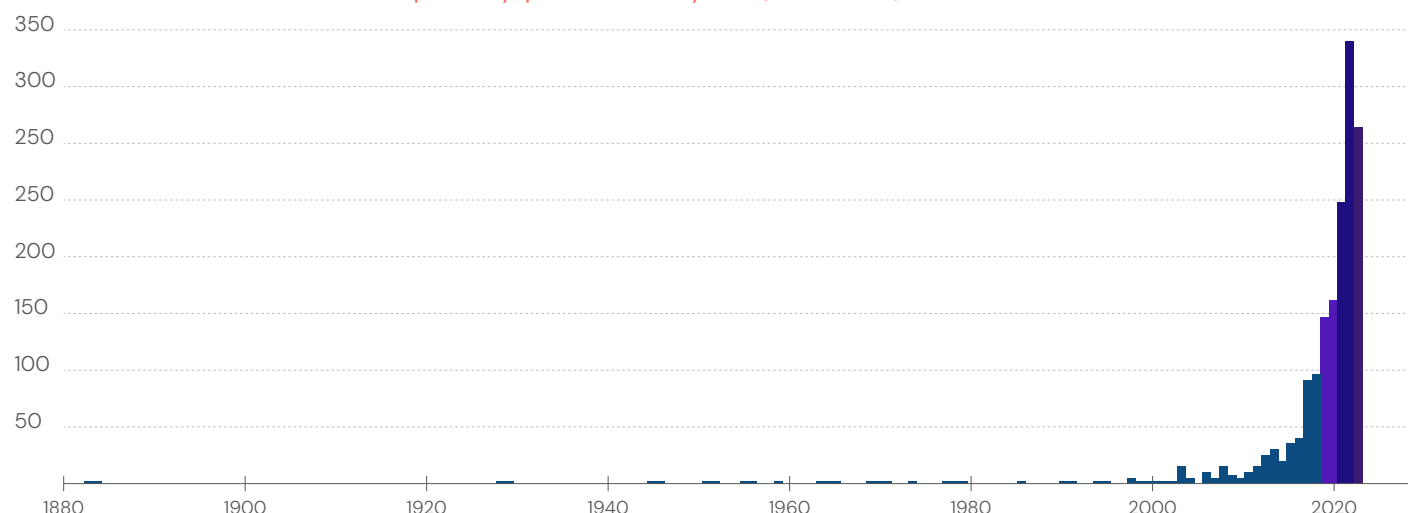
Note: Classification as 'low' or 'medium' did not mean automatic exclusion. Instead, these labels triggered further review of the source and a judgment about whether and in what context a source should be cited.

4 Characteristics of the Cited Sources

The sources were collected in Zotero. The evidence base contains 3,095 entries, of which 1,664 sources are cited in this report. Of these, 97% were published after 2000 and 90% after 2015. Figures A1 to A3 provide data on the number and share of citations in the report by publication year.

Figure A1

Number of citations in the report by publication year (N = 1,664)



Note: The gradient colors in the chart represent the number of cited publications in each year. The year with the highest number of cited publications is in dark purple, the year with the lowest number is in blue, and bars in between are shaded from the lowest number – light blue to highest number – darkest purple.

Figure A2

Number of citations in the report
by publication year (2000–2014, N = 110)

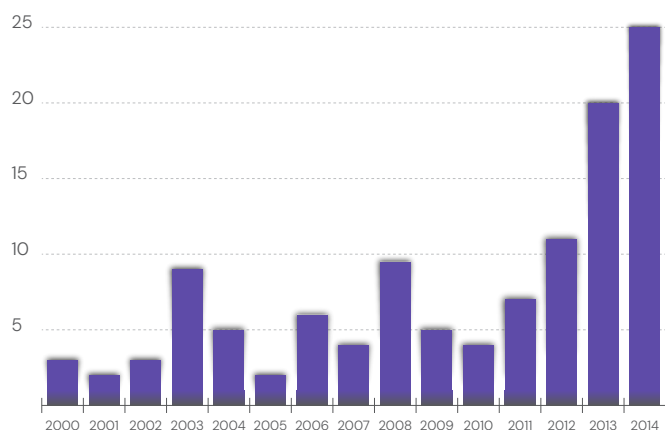
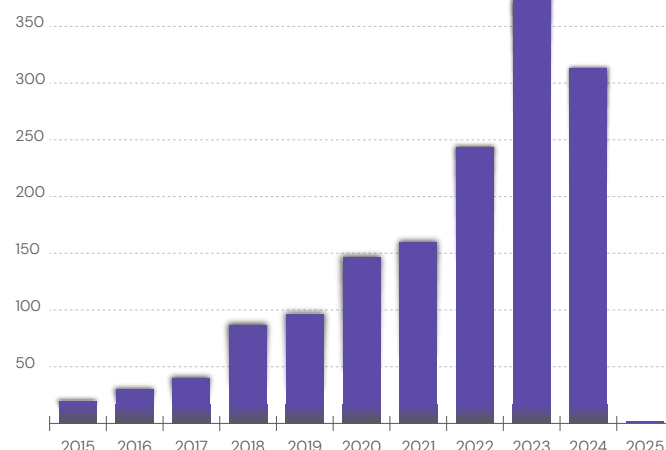


Figure A3

Number of citations in the report by
publication year (2015–2024, N = 1,508)



Of the cited sources, journal articles were the most common publication type, followed by reports, and then books: 43% of all the 1,664 citations were peer-reviewed articles, 20% were books and book chapters, 18% were reports, and 18% were 'other materials', including conference papers, newspaper and magazine articles, and blog posts. A decision was taken to target the collection of sources principally from 2022 to 2024 due to the volume of sources available over a longer time period and the Observatory's emphasis on current information ecosystems issues. This means that the sources were more diverse in recent years, especially after 2015. Figures A4 to A7 show these distributions.

Figure A4

Most common publication types
(N=1,664)

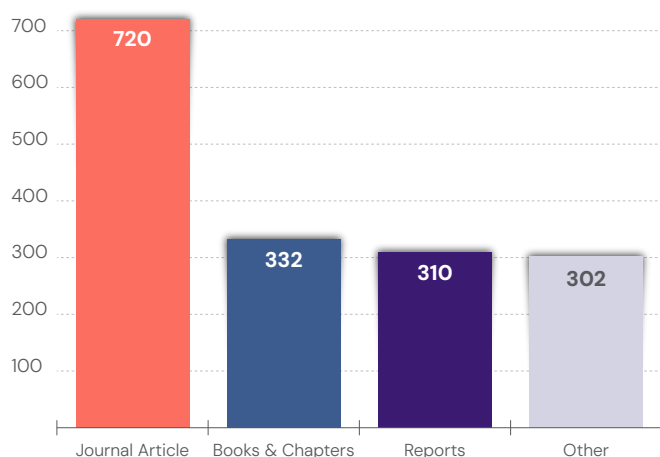


Figure A5

Number of publication types by year,
2000–2024 (N = 1,618)

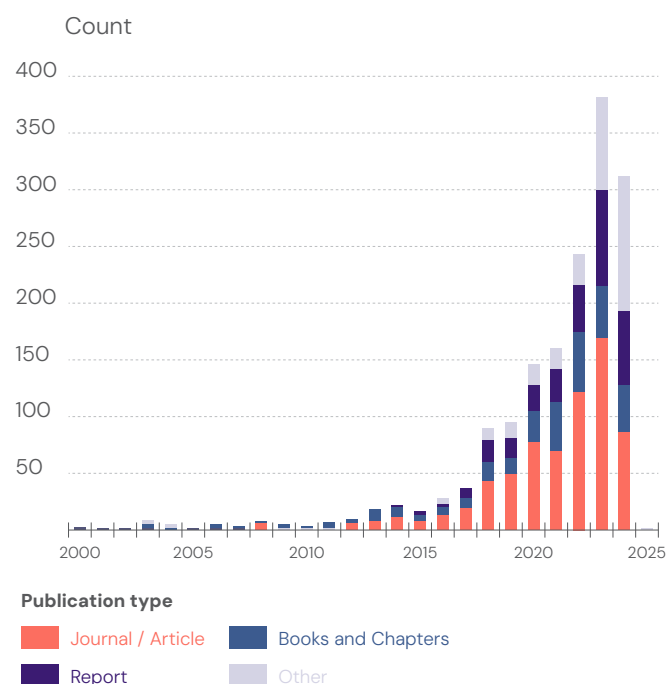


Figure A6

Types by year, bar chart, 2000–2024
(N = 1,618)

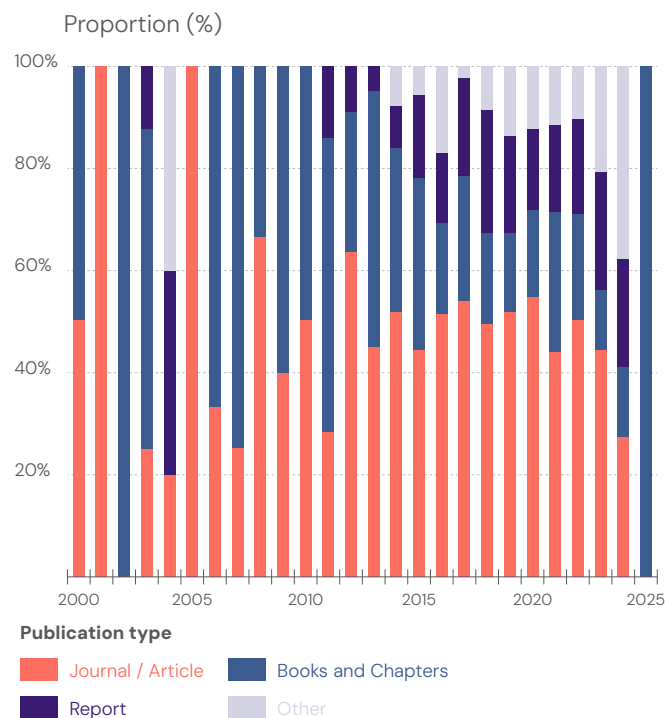
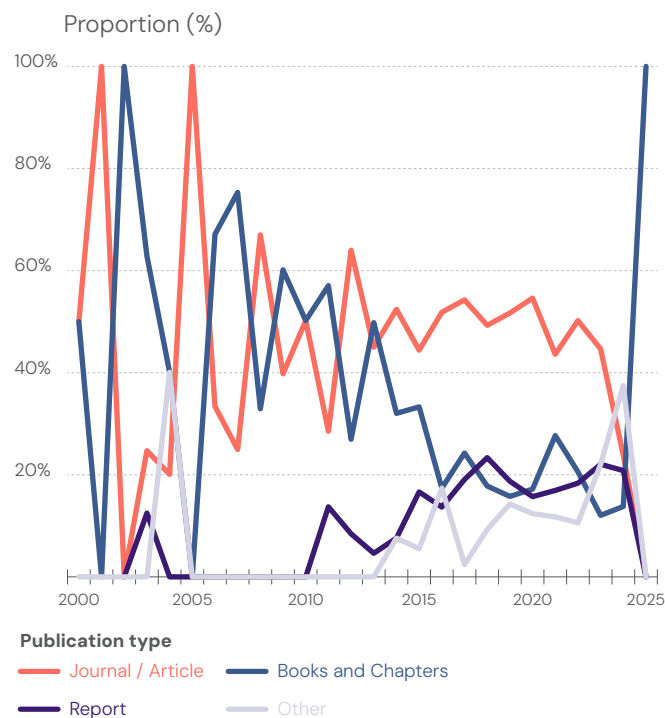


Figure A7

Types by year, cumulative, 2000–2024
(N = 1,618)



A regional breakdown of the cited sources shows that they are skewed towards literature from the Global North, making up 65.5% (N=1,664). Despite our effort to address this imbalance (see Section 2.2), and several interviews with experts in the Global Majority World, only 22.5% are either about the Global Majority World or by authors from the Global Majority World. The remaining 12% are classified as Global – meaning that they draw on arguments or evidence from a reasonably large number of countries in both the Global North and Global Majority World.

The proportion of Global Majority World citations in the report increased from May 2023 (the first draft report stage) to October 2024, when the source collection process was closed (15% in May 2024 to 22.5% in October 2024). This signals the structural bias of research towards the Global North, and it means that the evidence in this report relies disproportionately on Global North sources. The absence of a robust range of perspectives reflecting the experience of information ecosystems in the Global Majority World is noted throughout the report. Figures A8 to A10 visualize the distribution of sources cited in the report.

Figure A8

Number of types of publications, by regional focus (N = 1,664)

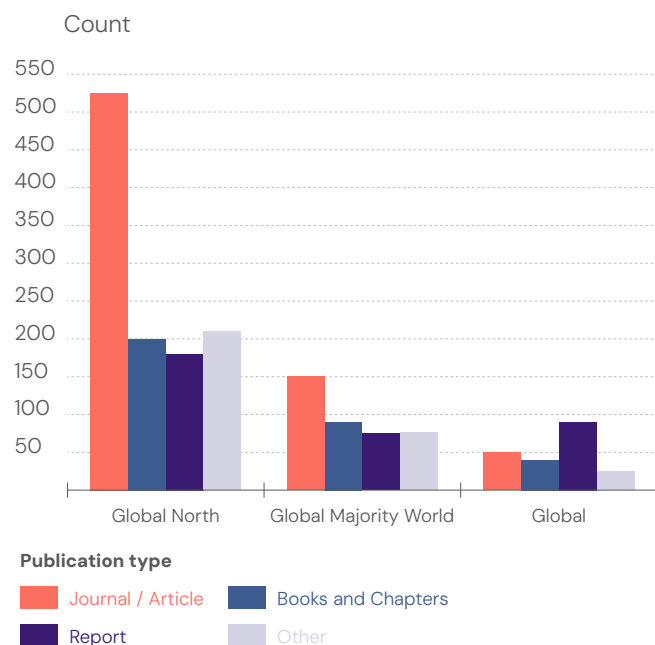


Figure A9

Share of types of publications, by regional focus (N = 1,664)

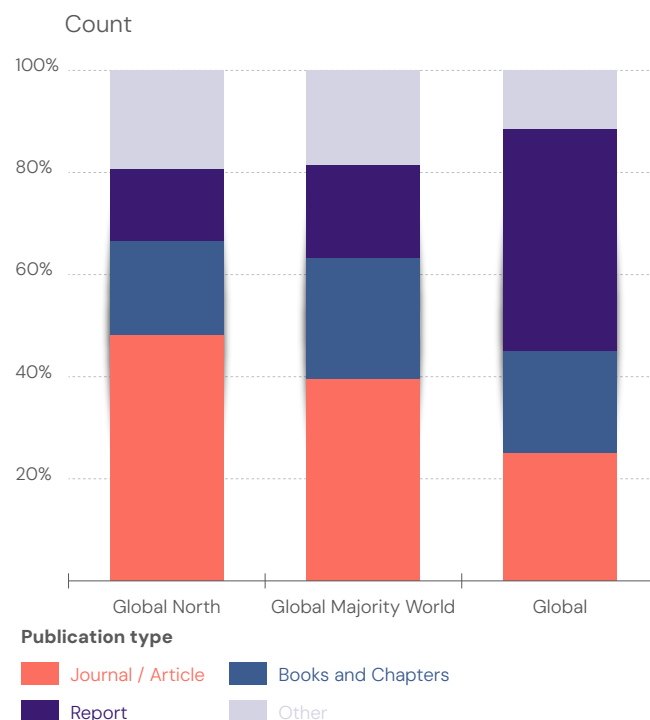
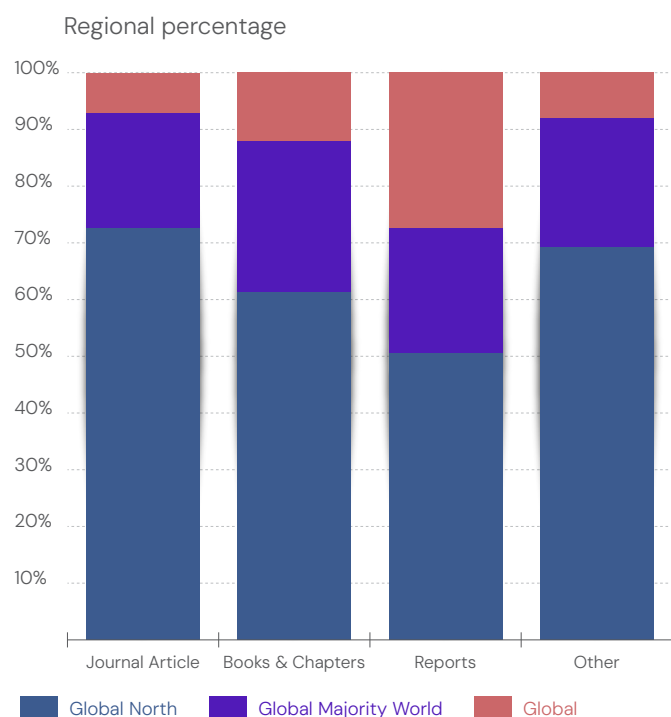


Figure A10

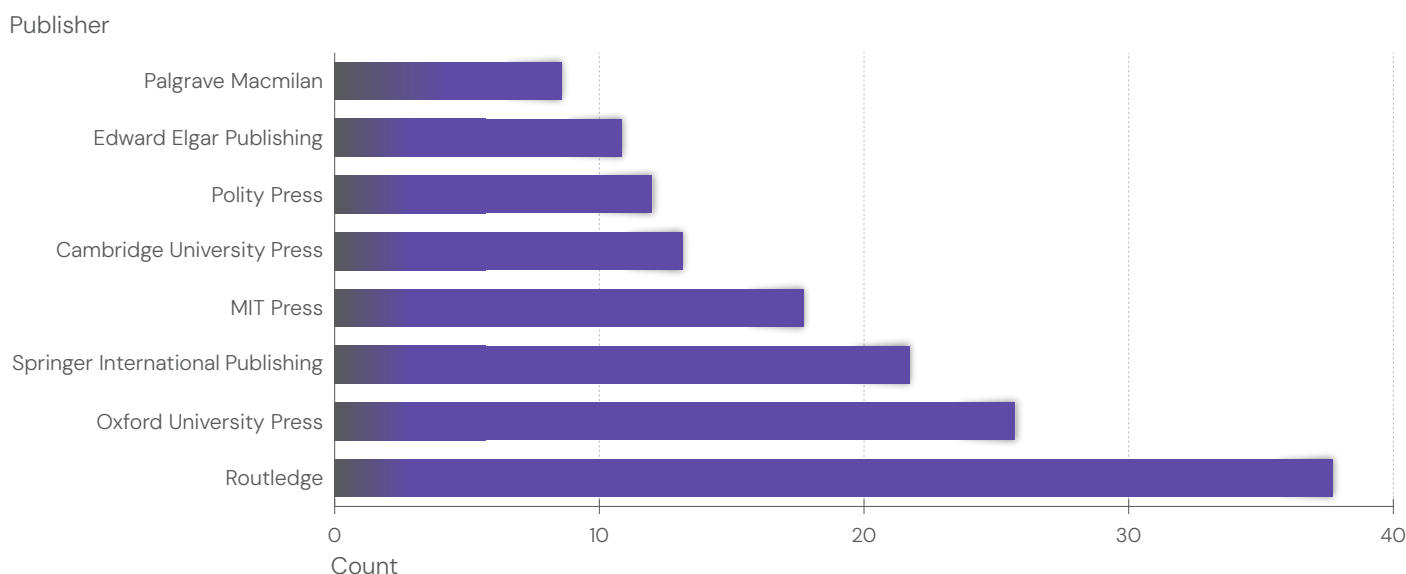
Share of types of cited publications by regional focus (N=1,664)



The publishers of books and book chapters came from diverse sources, most of which are based in the United States, the United Kingdom or Western Europe and often affiliated with academic institutions (e.g., Stanford University, Duke University, University of Amsterdam). In total, we cite 332 books and book chapters from 133 different publishing companies. International academic publishing companies were the most frequently cited (see Figure 11), with the top 20 publishers accounting for 14% of all cited sources (N=1,664) and 37% of all books and book chapters (N=332). Among these, Routledge, Oxford University Press, Springer International Publishing were the predominant publishers, accounting for 10%, 8% and 6%, respectively, of the 332 cited books and book chapters.

Figure A11

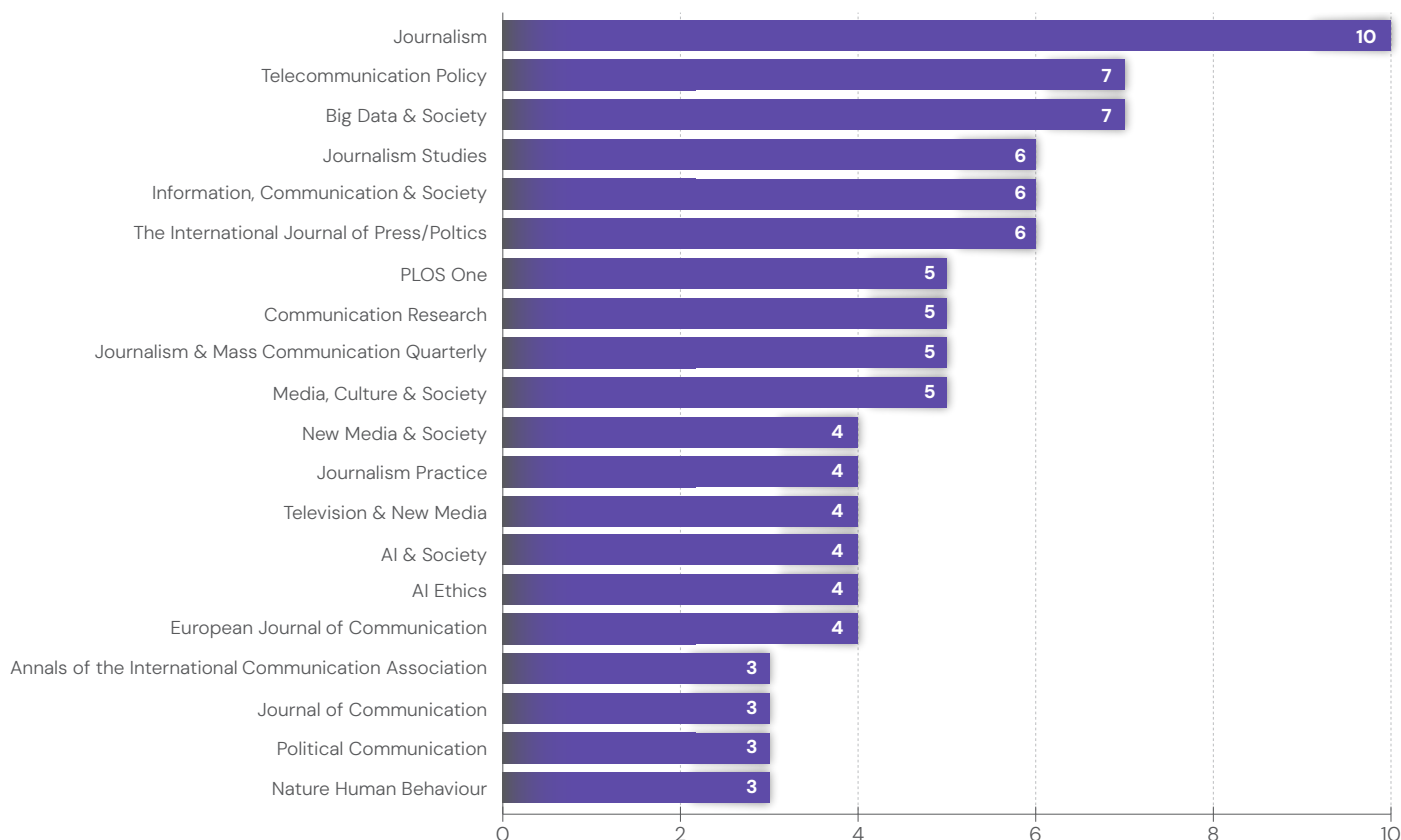
Number of books and book chapters by most frequently occurring publishing company



Finally, the report cites a total of 712 articles from 132 different journals. Among these, the most cited were in the media, journalism and communications fields, with *Journalism*, *Telecommunications Policy* and *Big Data & Society* being the most commonly cited. Figure A12 provides a summary.

Figure A12

Most frequently cited journals in the report



5 Concepts and Definitions

The terminology used in this report was informed by scholarly considerations, with the aim of making the report visible in policy communities. We used the terms ‘information ecosystems’ and ‘healthy information ecosystems’ as well as ‘public sphere’, and this section discusses some of the considerations that led to these choices.

In the late 1990s, information ‘ecologies’ were defined in the scholarly literature as systems of ‘people, practices, values, and technologies in a particular local environment’.⁸ Research on the ‘health’ of information ‘ecologies’ was undertaken focusing on relationships among people, information and technology in contexts such as libraries and other local communities. In this context, health was examined in relation to values and how these changed with the priorities of specific stakeholders. A ‘healthy information ecosystem’ has been defined more recently as a ‘balanced and well-functioning system of information creation, exchange, flow and utilization’, with components organized in non-hierarchical and non-linear ways. Building partly on the tradition of ‘ecology’ and ‘ecosystem’ studies, this approach puts the ‘media system’ at the center and locates practices such as content labeling and classification, data, content moderation, authentication, digital platforms and business models as well as data, AI systems including machine learning on the infrastructure layer of an information ecosystem.⁹ Additionally, the ‘health’ of ‘information ecosystems’ has been investigated, with ‘ecosystem’ defined as ‘a complex network or system of interacting organisms and their physical environment. An ecosystem is similarly characterized by ‘the interdependent relationships among its components’,¹⁰ and an effort is made to include cultural, social, political and technological components. This work has focused on ‘healthy values’, for example diversity, equity, inclusion and accessibility, and is often framed by strands of complex systems theory. In this context, information ecosystems are understood as evolutionary, indeterminate and self-organizing. If power asymmetries among actors in the ecosystem are mentioned, this is typically undertheorized due to the emphasis on the indeterminant, that is, unpredictable, system changes.

‘Information ecosystems’ terminology is increasingly present in policy documents and in a variety of research traditions.¹¹ In other research traditions, however, this metaphor is strongly criticized. It is said, for example, to obfuscate the requirements for democratic governance because it serves as ‘a means of justification and legitimacy under contemporary neoliberalized orders that typically chafe at modes of public intervention and the language of democratic statecraft’.¹² In this sense it is argued that the metaphor distracts attention from investigation of whose interests are served by big tech company actions, and ‘by whom, and by what right, can someone be excluded from public conversation?’¹³ Further, it is argued that it makes no distinction ‘between trolling, intent to harm, and justified outrage’.¹⁴ The metaphor of an ecosystem is understood by some critics to de-emphasize human agency because ‘system’ concepts are assumed to operate like natural systems.

⁸ Nardi & O’Day (1999, p. 49).

⁹ Radsch (2023e, p. 2–3).

¹⁰ Introne *et al.* (2024, p. 1030).

¹¹ This metaphor has been used recently by the Center for Democracy and Technology (CDT), US, International Federation of Library Associations and Institutions (IFLA), Open Knowledge Foundation, Centre for Media Pluralism and Freedom (CMPF), Research ICT Africa (RIA), United Nations (UN), United Nations High Commissioner for Refugees (UNHCR), Organisation for Economic Co-operation and Development (OECD), Partnership on AI, as well as numerous law firms, plus Meta/Facebook among other companies. ‘Information ecosystem’ in the scholarly literature appears infrequently up to 1990, except in relation to the environment and education. From 1990–2000 it appears in relation to multimedia; from 2000–10 it is mainly used in relation to mobile communication and 4G, in business studies and in relation to the World Wide Web and multimodal communication. From 2010–20, it is present in relation to 5G, media and media convergence, social media and journalism, as well as media literacy and technology. Most recently, it appears in relation to medical health and the Covid-19 pandemic, government uses of digital systems, AI systems, ‘fake news’, social media and echo chambers. As of October 2024, Google Scholar has yielded about 14,200 entries for ‘information ecosystem’ and 3,580 for ‘information ecosystems’. ‘Healthy information ecosystems’ appears 43 times and 125 times for ‘healthy information ecosystem’. Most occurrences are in papers from the Global North, but also in publications in the Global Majority World. The metaphor of an ecology or ecosystem of media or communication is used extensively in the scholarly literature in the social sciences and sciences, but this string was not reviewed.

¹² Gibson *et al.* (2023, p.2). One author was a senior researcher at Microsoft Research.

¹³ Gibson *et al.* (2023, p. 11).

¹⁴ Rieder & Skop (2021, p. 12), supported by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft).

Complex systems theory is said to neglect ‘deliberative conversations and subaltern counterpublics’.¹⁵ The phrase ‘healthy information ecosystem’ is also said to serve as ‘a useful tactic for journalists and policymakers, who do not recognize or may wish to cloak their own normative position’.¹⁶ For all these reasons, the healthy information ecosystems metaphor is positioned as serving the interests of big tech companies, shielding them from responsibility for creating problems.¹⁷

In some of the literature on information ecosystems the historical specificity and the role of imaginaries in structuring the public sphere is neglected.¹⁸ In this context, many scholars prefer to focus on the ‘public sphere’, a sphere or space in which, ideally, public deliberation can proceed rationally.¹⁹ This concept is also subject to criticism in some research traditions for its assumptions about the existence of ‘critical-rational publics’.²⁰ The public sphere concept is criticized for its assumptions about liberalism and illiberalism,²¹ and its emphasis on ideal speech conditions and rational communicative action. Historical and contemporary exclusions and marginalizations from the public sphere are neglected in some writing about the role of the public sphere in society.²² The concept is also said to neglect the importance of distinguishing between the normative conditions for a democratic polity where participants struggle to secure the rights to which they are entitled, and the Eurocentricity of the ‘public sphere’ concept is criticized often, but not exclusively, by scholars in the Global Majority World.²³ The concept is criticized for neglecting the conflictual nature of political discourse and the role of emotions in debate in the public sphere.²⁴ Finally, less attention tends to be given to structural power, ideology and cultural specificity in studies of the public sphere than is the case, for example, in studies that aim to explain how power asymmetries are embedded and maintained in information and communication systems.

In this report, we use the ‘information ecosystems’ terminology but do not conceive of their health as a pathology in the medical sense. Where the health of information ecosystems is discussed, it is in relation to normative choices about values; specifically, those enshrined in international human rights commitments and in relation to private and public institutional capacities to uphold those values. The metaphor of an ‘information ecosystem’ does tend in much of the literature to be only loosely coupled with values-based considerations and with power dynamics that may have uncertain outcomes, but which also constrain actions to reduce power asymmetries in the short and medium term. For this reason, we draw on theories in the political economy tradition, but not to suggest that political economy structures are determining of certain outcomes. Instead we draw on these theories to assess the scope for the exercise of counter-power and individual and collective agency.

We also focus on the ‘public sphere’ to center communicative practices that are associated with contemporary problems facing democracy at the core of our analysis.²⁵ We position the ‘public sphere’ centrally in our conception of information ecosystems because:

¹⁵ Ehrenfeld (2020, p. 308); Mouffe (1966).

¹⁶ Gibson *et al.* (2023, p. 12).

¹⁷ Gibson *et al.* (2023, p. 10); Ehrenfeld (2020, p. 308).

¹⁸ Ehrenfeld (2020). The ‘public sphere’ appears throughout the academic literature, although less frequently in policy studies (apart from media and political science studies). As of October 2024, Google Scholar reported approximately 1.8 million mentions, 17,700 since the beginning of 2024, with plural usage ‘spheres’ occurring 130,000 times. The concept ‘public sphere’ appears, e.g., in relation to sphericules, postmodernism, the post-public sphere, citizenship and democracy, modernity, deliberation, culture, radicalization and the internet.

¹⁹ Habermas (2015), first published in English in 1989, in German in 1962; Habermas (2022); see also Fraser (1992).

²⁰ Gerbaudo (2022).

²¹ Štětka & Mihelj (2024a, p. 31).

²² Dahlberg (2014); Devenney (2009); Habermas (2022).

²³ See Banaji (2024); de Sousa Santos (2018).

²⁴ Cammaerts (2024, p. 27).

²⁵ Cammaerts (2024); Splichal (2022b, p. 213).

Without a model of relationality that attunes us to the historically distinct nature of the public sphere, we lack the means to reflect about how our own acts might contribute to the reinforcement of new forms of solidarity, new forms of intimacy, new forms of collective action, and new forms of identification – in short, new forms of being “public” with strangers.²⁶

On both the individual and institutional levels of analysis a key interest in this research assessment is with power relationships and struggles among actors over the design and operation of ‘information ecosystems’ and the characteristics of the spaces they create for public dialogue.

²⁶ Ehrenfeld (2020, p. 308).

BIBLIOGRAPHY

- 2020 Election Research Project. (2020). A proposal for understanding social media's impact on elections: Peer-reviewed scientific research. *Medium*, 31 August. https://medium.com/@2020_election_research_project/a-proposal-for-understanding-social-medias-impact-on-elections-4ca5b7aae10
- Aalberg, T., Blekesaune, A., & Elvestad, E. (2013). Media choice and informed democracy: Toward increasing news consumption gaps in Europe? *The International Journal of Press/Politics*, 18(3), 281–303. <https://doi.org/10.1177/1940161213485990>
- Aaronson, S. A. (2021). Data is Different, Policymakers Should Pay Attention to Its Governance. In M. Burri (Ed.), *Big Data and Global Trade Law* (pp. 340–360). Cambridge University Press.
- Aaronson, S. A., & Leblond, P. (2018). Another digital divide: The rise of data realms and its implications for the WTO. *Journal of International Economic Law*, 21(2), 245–272. <https://doi.org/10.1093/jiel/jgy019>
- Abdalla, M., & Abdalla, M. (2021). The Grey Hoodie Project: Big tobacco, big tech, and the threat on academic integrity. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. <https://doi.org/10.1145/3467102.3462563>
- Abdulrauf, L. A., & Dube, H. (Eds) (2024). *Data Privacy Law in Africa: Emerging Perspectives*. Pretoria University Law Press.
- Acemoglu, D., & Johnson, S. (2023). *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. Basic Books.
- Adams, R. (2022). *AI in Africa: Key Concerns and Policy Considerations for the Future of the Continent*. Africa Policy Research Institute (APRI), Germany, Policy Brief.
- Adel, K., Elhakeem, A., & Marzouk, M. (2023). Decentralized system for construction projects data management using blockchain and IPFS. *Journal of Civil and Engineering Management*, 29(4), 342–359. <https://doi.org/10.3846/jcem.2023.18646>
- Adjin-Tettey, T. D. (2022). Combating fake news, disinformation, and misinformation: Experimental evidence for media literacy education. *Cogent Arts & Humanities*, 9(1), 1–17. <https://doi.org/10.1080/23311983.2022.2037229>
- Adobe, Amazon, Anthropic, arm, et al. (2024). *A Tech Accord to Combat Deceptive Use of AI in 2024 Elections*. www.aielectionsaccord.com
- ADNRN (African Digital Rights Network). (2024). *The Digitalisation of Social Protection in Africa: The Perspective of Marginalised Workers*. www.africandigitalrightsnetwork.org/publications
- Advox Team (2024). *The Unfreedom Monitor: A Methodology for Tracking Digital Authoritarianism Around the World*. Global Voices Advox, DW Akademie and German Cooperation (Deutsche Zusammenarbeit). https://globalvoices.org/wp-content/uploads/2023/08/Unfreedom_Monitor_Observatory_Final_Report_2023.pdf
- Afina, Y., Buchser, M., Krasodomski, A., Rowe, J., Sun, N., & Wilkinson, R. (2024). *Towards a Global Approach to Digital Platform Regulation: Preserving Openness Amid the Push for Internet Sovereignty*. Chatham House, Royal Institute of International Affairs.
- Africa Center for Strategic Studies. (2024). *Mapping a Surge of Disinformation in Africa*. <https://africacenter.org/spotlight/mapping-a-surge-of-disinformation-in-africa>
- African Union. (2014). African Union Convention on Cyber Security and Personal Data Protection. <https://africanlii.org/akn/aa-au/act/convention/2014/cyber-security-and-personal-data-protection/eng@2014-06-27>
- African Union. (2024). Continental Artificial Intelligence Strategy: Harnessing AI for Africa's development and prosperity. African Union Press Release, 9 August. <https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy#:~:text=The%20Continental%20AI%20Strategy%20calls,inclusive%20and%20responsible%20AI%20development>
- Aguerre, C., Campbell-Verduyn, M., & Scholte, J. A. (Eds) (2024). *Global Digital Data Governance: Polycentric Perspectives*. Routledge.
- Aharoni, T., Kligler-Vilenchik, N., & Tenenboim-Weinblatt, K. (2021). 'Be less of a slave to the news': A text-material perspective on news avoidance among young adults. *Journalism Studies*, 22(1), 42–59. <https://doi.org/10.1080/1461670X.2020.1852885>
- Ahmad, S., & Greb, M. (2022). Automating social media content moderation: Implications for governance and labour discretion. *Work in the Global Economy*, 2(2), 176–198. <https://doi.org/10.1332/273241721X16647876031174>
- Ahmed, N., Wahed, M., & Thompson, N. C. (2023). The growing influence of industry in AI research. *Science*, 379(6635), 884–886. <https://doi.org/10.1126/science.ade2420>
- AI Elections Accord. (2024). Progress update, September. www.aielectionsaccord.com/progress-update
- AI Now Institute. (2024). *Redirecting Europe's AI Industrial Policy: From Competitiveness to Public Interest*. <https://ainowinstitute.org/redirecting-europes-ai-industrial-policy>
- Aimeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. *Social Network Analysis and Mining*, 13(30), 1–36. <https://doi.org/10.1007/s13278-023-01028-5>
- Ajuntament de Barcelona [Barcelona City Council]. (2023). *Definició de metodologies de treball i protocols per a la implementació de sistemes algorítmics*.
- Ajunwa, I., Crawford, K., & Schultz, J. (2017). *Limitless worker surveillance*. *Berkely Law*, 105, 735–776. <https://doi.org/10.15779/Z38BR8MF94>

- Akimenko, V., & Giles, K. (2020). Russia's cyber and information warfare. *Asia Policy*, 15(2), 67–75. www.jstor.org/stable/27023902
- Akman, P. (2022). Regulating competition in digital platform markets: A critical assessment of the framework and approach of the EU Digital Markets Act. *European Law Review*, 47(2022), 85–114. <https://dialnet.unirioja.es/servlet/articulo?codigo=8320879>
- Akser, M., & Baybars, B. (2023). Repressed media and illiberal politics in Turkey: The persistence of fear. *Southeast European and Black Sea Studies*, 23(1), 159–177. <https://doi.org/10.1080/14683857.2022.2088647>
- Alaimo, C., & Kallinikos, J. (2024). *Data Rules: Reinventing the Market Economy*. MIT Press.
- Alexander, F. (2024). UN attempts AI power grab. The West is unhappy. CEPA, 24 July. <https://cepa.org/article/un-attempts-ai-power-grab-the-west-is-unhappy>
- Alfonsi, C. (2019). Taming tech giants requires fixing the revolving door. *Kennedy School Review*, 19(2019), 166–170. www.proquest.com/openview/cea84d87b730cle8e8db055e84b5c9c1/1?pq-origsite=gscholar&cbl=436382
- Alieva, I., Moffitt, J. D., & Carley, K. M. (2022). How disinformation operations against Russian opposition leader Alexei Navalny influence the international audience on Twitter. *Social Network Analysis and Mining*, 12(80), 1–13. <https://doi.org/10.1007/s13278-022-00908-6>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 32(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Allen, D., & Weyl, E. G. (2024). The real dangers of generative AI. *Journal of Democracy*, 35(1), 146–162. www.journalofdemocracy.org/articles/the-real-dangers-of-generative-ai
- Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, 6(14), 1–6. <https://doi.org/10.1126/sciadv.aay3539>
- Altay, S., & Acerbi, A. (2023). People believe misinformation is a threat because they assume others are gullible. *New Media & Society*, 26(11), 6440–6461. <https://doi.org/10.1177/14614448231153379>
- Altay, S., Berriche, M., & Acerbi, A. (2023a). Misinformation on misinformation: Conceptual and methodological challenges. *Social Media + Society*, 9(1), 1–13. <https://journals.sagepub.com/doi/epub/10.1177/20563051221150412>
- Altay, S., Fletcher, R., & Nielsen, R. K. (2024). News participation is declining: Evidence from 46 countries between 2015 and 2022. *New Media & Society, Online First*, 1–22. <https://doi.org/10.1177/14614448241247822>
- Altay, S., Nielsen, R. K., & Fletcher, R. (2023b). News can help! The impact of news media and digital platforms on awareness of and belief in misinformation. *The International Journal of Press/Politics*, 29(2), 459–484. <https://doi.org/10.1177/19401612221148981>
- Alufaisan, Y., Marusich, L. R., Bakdash, J. Z., Zhou, Y., & Kantarcioglu, M. (2021). Does explainable artificial intelligence improve human decision-making? *Proceeding of the AAAI Conference on Artificial Intelligence*, 35(8), 6618–6626. <https://doi.org/10.1609/aaai.v35i8.16819>
- American Sunlight Project. (2024). *The Information Laundering Cycle: How a Coordinated Effort Weaponized the American Political System in Favor of Disinformation*. www.americansunlight.org/ilc-database
- Amnesty International, & AI Forensics. (2023). *Driven into the Darkness: How TikTok's 'For You' Feed Encourages Self-Harm and Suicidal Ideation*. www.amnesty.org/en/documents/POL40/7350/2023/en
- Ananny, M. (2023). What a 'platform press' view has to offer. *Digital Journalism*, 11(8), 1568–1575. <https://doi.org/10.1080/21670811.2023.2257768>
- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. <https://doi.org/10.1177/1461444816676645>
- Andere, B., & Kathure, M. (2024). *Strengthening Data Protection in Africa: Key Issues for Implementation*. Access Now. www.accessnow.org/wp-content/uploads/2024/01/Strengthening-data-protection-in-Africa-key-issues-for-implementation-updated.pdf
- Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Knopf Doubleday Publishing Group.
- Anderson, C. W. (2021). Fake news is not a virus: On platforms and their effects. *Communication Theory*, 31(1), 42–61. <https://doi.org/10.1093/ct/qtaa008>
- Anderson, N. (2024). After cybersecurity lab wouldn't use AV software, US accuses Georgia Tech of fraud. *Ars Technica*, 23 August. <https://arstechnica.com/security/2024/08/oh-your-cybersecurity-researchers-wont-use-antivirus-tools-heres-a-federal-lawsuit>
- Ang, P.-H., & Haristya, S. (2024). The governance, legitimacy and efficacy of Facebook's Oversight Board: A model for global tech platforms? *Emerging Media*, 2(2). <https://doi.org/10.1177/27523543241266860>
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. ProPublica, 23 May. www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- Annan, K. (1997). If information and knowledge are central to democracy, they are conditions for development. United Nations Press Release SG/SM/6268, 23 June. <https://press.un.org/en/1997/19970623.sgsm6268.html>
- Annoni, A., Benczur, P., Bertoldi, P., Delipetrev, B., et al. (2018). *Artificial Intelligence: A European Perspective*. European Commission, JRC EUR 29425.

- Anthony, A., Sharma, L., & Noor, E. (2024). Advancing a More Global Agenda for Trustworthy Artificial Intelligence. Carnegie Endowment for International Peace, 30 April. <https://carnegieendowment.org/research/2024/04/advancing-a-more-global-agenda-for-trustworthy-artificial-intelligence?lang=en>
- Arcom. (2024). *The French and Information (Les Français et information)*. www.arcom.fr/nos-ressources/etudes-et-donnees/etudes-bilans-et-rapports-de-larcom/les-francais-et-linformation
- Arendt, H. ([1958] 1998). *The Human Condition*. University of Chicago Press.
- Arendt, H. (1968). Truth and Politics. In H. Arendt, *Between Past and Future: Eight Exercises in Political Thought* (pp. 227–264). Viking Press.
- Arriagada, A., Bonhomme, M., Ibanez, F., & Leyton, J. (2023). The gig economy in Chile: Examining labor conditions and the nature of gig work in a Global South country. *Digital Geography and Society*, 5(2023), 1–6. <https://doi.org/10.1016/j.diggeo.2023.100063>
- Arguete, N., & Calvo, E. (2023). *Nosotros contra ellos: Cómo trabajan las redes para confirmar nuestras creencias y rechazar las de los otros [Us against them: How networks work to confirm our beliefs and reject those of others]*. Sociologie y Política.
- Ashok, P. (2023). The curious case of automated decision-making in India. *International Cybersecurity Law Review*, 4(2), 235–248. <https://doi.org/10.1365/s43439-022-00078-3>
- Aufderheide, P. (1993). *Media Literacy: A Report of the National Leadership Conference on Media Literacy*. Aspen Institute Communications and Society Program.
- Aula, V. (2023). Evidence-based policymaking in the legislatures: Timeliness and politics of evidence in Finland. *Policy & Politics*, 51(4), 673–694. <https://doi.org/10.1332/O30557321X16873376680333>
- Avle, S. (2022). Hardware and data in the platform era: Chinese smartphones in Africa. *Media, Culture & Society*, 44(8), 1473–1489. <https://doi.org/10.1177/01634437221128935>
- AWS (no date) Just Walk Out technology. <https://aws.amazon.com/just-walk-out>
- Bachelet, M. (2019). Human rights in the digital age – Can they make a difference? OHCHR Keynote Speech. www.ohchr.org/en/speeches/2019/10/human-rights-digital-age
- Baecker, C., Alabbadi, O., Yogiputra, G. P., & Tien Dung, N. (2023). Threats Provided by Artificial Intelligence that Could Disrupt the Democratic System. Scientific Paper, University of Applied Science, Brandenburg.
- Bagchi, C., Menczer, F., Lundquist, J., Tarafdar, M., Paik, A., & Grabowicz, P. (2024). Social media algorithms can curb misinformation, but do they? *ArXiv*. <https://doi.org/10.48550/arXiv.2409.18393>
- Bail, C. (2021). *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing*. Princeton University Press.
- Bailey, R., Parsheera, S., & Sane, R. (2020). Comments on the ‘Report by the Committee of Experts on Non-Personal Data Governance Framework’ to Ministry of Electronics and IT, Government of India. <https://doi.org/10.2139/ssrn.3724184>
- Baines, D., & Elliott, R. (2020). *Defining Misinformation, Disinformation and Malinformation: An Urgent Need for Clarity During the COVID-19 Infodemic*. Discussion Papers 20–16, Department of Economics, University of Birmingham.
- Bakir, V., & McStay, A. (2018). Fake news and the economy of emotions: Problems, causes, solutions. *Digital Journalism*, 6(2), 154–175. <https://doi.org/10.1080/21670811.2017.1345645>
- Bakir, V., & McStay, A. (2022). *Core Incubators of False Information Online*. In V. Bakir & A. McStay, *Optimising Emotions, Incubating Falsehoods: How to Protect the Global Civic Body from Disinformation and Misinformation* (pp. 29–52). Springer International Publishing.
- Bakke, N. A., & Barland, J. (2022). Disruptive innovations and paradigm shifts in journalism as a business: From advertisers first to readers first and traditional operational models to the AI factory. *SAGE Open*, 12(2), 1–18. <https://doi.org/10.1177/21582440221094819>
- Bakker, B. N., & Lelkes, Y. (2024). Putting the affect into affective polarisation. *Cognition and Emotion*, 38(4), 418–436. <https://doi.org/10.1080/02699931.2024.2362366>
- Balkin, J. M. (2017). Free speech in the algorithmic society: Big data, private governance, and new school speech regulation. *UC Davis Law Review*, 51, 1151–1209. <https://lawreview.law.ucdavis.edu/archives/51/3/free-speech-algorithmic-society-big-data-private-governance-and-new-school-speech>
- Ball, K., & Snider, L. (2013). *The Surveillance-Industrial Complex: A Political Economy of Surveillance*. Routledge.
- Banaji, S. (2024). Against resilience: The (anti-)ethics of participation in an unjust and unequal public sphere. *Javnost – The Public*, 31(1), 141–157. <https://doi.org/10.1080/13183222.2024.2311012>
- Banaji, S., & Bhat, R. (2022). *Social Media and Hate*. Routledge.
- Banaji, S., Bhat, R., Agarwal, A., Passanha, N., & Pravin, M. S. (2019). *WhatsApp Vigilantes: An Exploration of Citizen Reception and Circulation of WhatsApp Misinformation Linked to Mob Violence in India*. Department of Media and Communications, London School of Economics and Political Science.
- Banerjee, S., Mont’Alverne, C., Ross Arguedas, A., Toff, B., Fletcher, R., & Nielsen, R. K. (2023). *Strategies for Building Trust in News: What the Public Say They Want Across Four Countries* Reuters Institute for the Study of Journalism, University of Oxford.
- Bannerman, S. (2022). Platform imperialism, communications law and relational sovereignty. *New Media & Society*, 26(4), 1816–1833. <https://doi.org/10.1177/14614448221077284>

- Bannerman, S., McNeil, B., Wyndham-West, K., & Beny, H. (2024). *2023 TechLobby Annual Report*. The Tech Lobby. <https://thetechlobby.ca/2023-techlobby-annual-report-interactive-graphics>
- Bansal, V. (2024). This delivery app takes away health insurance when workers don't meet quotas. Rest of World Reporting Global Tech Stories, 12 April. <https://restofworld.org/2024/swiggy-health-insurance-quotas>
- Baptista, J. P., & Gradim, A. (2022). Online disinformation on Facebook: The spread of fake news during the Portuguese 2019 election. *Journal of Contemporary European Studies*, 30(2), 297–312. <https://doi.org/10.1080/14782804.2020.1843415>
- Baranes, E. (2014). The interplay between network investment and content quality: Implications to net neutrality on the Internet. *Information Economics and Policy*, 28, 57–69. <https://doi.org/10.1016/j.infoecopol.2014.07.002>
- Barba-Kay, A. (2023). *A Web of Our Own Making: The Nature of Digital Formation*. Columbia University Press.
- Bargaoanu, A., & Radu, L. (2018). Fake news or Disinformation 2.0? Some insights into Romanians' digital behaviour. *Romanian Journal of European Affairs*, 18(1), 24–38. <https://ssrn.com/abstract=3296922>
- Barocas, S., & Nissenbaum, H. (2014). Big Data's End Run Around Anonymity and Consent. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds), *Privacy, Big Data, and the Public Good* (pp. 44–75). Cambridge University Press.
- Baron, G. R. (2002). *Now Is Too Late: Survival in an Era of Instant News*. Financial Times, Prentice Hall.
- Bateman, J., & Jackson, D. (2024). *Looking Ahead: Generative AI – Countering Disinformation Effectively: An Evidence-Based Policy Guide*. Carnegie Endowment for International Peace.
- Bates, J., Kennedy, H., Perea, I. M., Oman, S., & Pinney, L. (2023). Socially meaningful transparency in data-based systems: Reflections and proposals from practice. *Journal of Documentation*, 80(1), 1–20. <https://doi.org/DOI:10.1108/jd-01-2023-0006>
- Bauer, J. M., & Knieps, G. (2018). Complementary innovation and network neutrality. *Telecommunications Policy*, 42(2), 172–183. <https://doi.org/10.1016/j.telpol.2017.11.006>
- Bayer, J., Katsirea, I., Batura, O., Holznagel, B., Hartmann, S., & Lubianiec, K. (2021). *The Fight Against Disinformation and the Right to Freedom of Expression*. Report for the European Parliament's Committee on Civil Liberties, Justice and Home Affairs. Universities of Münster, Sheffield, Leuphana, and Ecorys, Poland.
- BBC Media Action. (2021). *Enabling Media Markets to Work for Democracy: An International Fund for Public Interest Media*. International Fund for Public Interest Media (IFPIM).
- Beckett, C. (2019). *New Powers, New Responsibilities: A Global Survey of Journalism and Artificial Intelligence*. POLIS (Department of Media and Communications), London School of Economics and Political Science, UK and Google News Initiative.
- Beckett, C., & Yaseen, M. (2023). *Generating Change: A Global Survey of What News Organisations Are Doing with AI*. POLIS (Department of Media and Communications), London School of Economics and Political Science, UK and Google News Initiative.
- Beiguelman, G. (2021). *Políticas da imagem: vigilância e resistência na dadosfera*. Ubu Editora.
- Belenguer, L. (2022). AI bias: Exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. *AI and Ethics*, 2(4), 771–787. <https://doi.org/10.1007/s43681-022-00138-8>
- Bell, E., Owen, T., Brown, P. D., Hauka, C., & Rashidian, N. (2017). *The Platform Press: How Silicon Valley Reengineered Journalism*. Tow Center for Journalism, Columbia Journalism School, US.
- Bellan, R. (2024). Meta axed CrowdTangle, a tool for tracking disinformation. Critics claim its replacement has just '1% of the features'. *TechCrunch*, 15 August. <https://techcrunch.com/2024/08/15/meta-shut-down-crowdtangle-a-tool-for-tracking-disinformation-heres-how-its-replacement-compares>
- Belli, L., Almeida, C., Couto, N., Radu, R., Weber, R. H., & Brown, I. (2022). *Towards Meaningful and Interoperable Transparency for Digital Platform: 2022 Outcome of the UN IGF Coalition on Platform Responsibility*. IGF (Internet Governance Forum).
- Bello y Villarino, J.-M. (2023). Global standard-setting for artificial intelligence: Para-regulating international law for AI? *The Australian Year Book of International Law Online*. https://brill.com/view/journals/auso/41/1/article-p157_7.xml
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In ACM (Ed.), *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623).
- Benequista, N., Abbott, S., Rothman, P., & Mano, W. (Eds) (2019). *International Media Development: Historical Perspectives and New Frontiers*. Peter Lang Publishers.
- Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons.
- Benkler, Y. (2020). A Political Economy of the Origins of Asymmetric Propaganda in American Media. In S. Livingston & W. L. Bennett (Eds), *The Disinformation Age* (pp. 43–66). Cambridge University Press.
- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press.
- Bennett, C. J., & Raab, C. (2006). *The Governance of Privacy: Policy Instruments in a Global Perspective*. MIT Press.
- Bennett, W. L., & Kneuer, M. (2023). Communication and democratic erosion: The rise of illiberal public spheres. *European Journal of Communication*, 39(2), 1–20. <https://doi.org/10.1177/02673231231217378>

- Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2), 122–139. <https://doi.org/10.1177/0267323118760317>
- Bennett, W. L., & Livingston, S. (2023). A Brief History of the Disinformation Age: Information Wars and the Decline of Institutional Authority. In S. Salgado & S. Papathanassopoulos (Eds), *Streamlining Political Communication Concepts: Updates, Changes, Normalcies* (pp. 43–73). Springer International Publishing.
- Bennett, W. L., & Livingston, S. (Eds) (2020). *The Disinformation Age: Politics, Technology and Disruptive Communication in the United States*. Cambridge University Press.
- Benson, J. (2019). Deliberative democracy and the problem of tacit knowledge. *Politics, Philosophy & Economics*, 18(1), 76–97. <https://doi.org/10.1177/1470594X18782086>
- Benton, J. (2022). Facebook looks ready to divorce the news industry, and I doubt couples counseling will help. NiemanLab, 16 June. www.niemanlab.org/2022/06/facebook-looks-ready-to-divorce-the-news-industry-and-i-doubt-couples-counseling-will-help
- Berdou, E. (2011). *Organization in Open Source Communities: At the Crossroads of the Gift and Market Economies*. Routledge.
- BEREC (Body of European Regulators for Electronic Communications). (2024a). *BEREC Guidelines on the Implementation by National Regulators of European Net Neutrality Rules*. www.berec.europa.eu/en/document-categories/berec/regulatory-best-practices/guidelines/berec-guidelines-on-the-implementation-by-national-regulators-of-european-net-neutrality-rules
- BEREC. (2024b). What is zero-rating? www.berec.europa.eu/en/what-is-zero-rating
- Berger, C., Freihse, C., Kettemann, M. C., Mosene, K., & Hofmann, V. (2023a). The EU Elections 2024: How to build resilience against disinformation campaigns on social platforms. Upgrade Democracy, 28 August. <https://upgradedemocracy.de/en/impulse/the-eu-elections-2024-how-to-build-resilience-against-disinformation-campaigns-on-social-platforms>
- Berger, C., Freihse, C., Mosene, K., Kettemann, M. C., & Hofmann, V. (2023b). Threats to democracy: Climate misinformation and gendered disinformation. Upgrade Democracy, 26 July. <https://upgradedemocracy.de/en/impulse/climate-misinformation-and-gender-related-disinformation-responsibilities-of-civil-society-public-sector-and-the-media>
- Berger, G., Gillwald, A., Orembo, E., Diouf, D., & Garcia, J. M. (2023). *Platform Problems and Regulatory Solutions: Findings from a Comprehensive Review of Existing Studies and Investigations*. Research ICT Africa, South Africa, Independent Researchers and Centro Latam Digitale, Mexico.
- Berkeley Economic Review. (2022). Elon Musk image. 17 November.
- Bernardi, A. J. B., & de Moraes, J. A. (2021). Fascismo à brasileira? Análise de conteúdo dos discursos de Bolsonaro após o segundo turno das eleições presidenciais de 2018. *Política & Sociedade*, 20(48), 300–327. <https://doi.org/10.5007/2175-7984.2021.72401>
- Berthelot, A., Caron, E., Jay, M., & Lefèvre, L. (2024). Estimating the environmental impact of Generative-AI services using an LCA-based methodology. *31st CIRP Conference on Life Cycle Engineering*, 122, 707–712. <https://doi.org/10.1016/j.procir.2024.01.098>
- Beskow, D. M., & Carley, K. M. (2020). Characterization and Comparison of Russian and Chinese Disinformation Campaigns. In K. Shu, S. Wang, D. Lee, & H. Liu (Ed.), *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities* (pp. 63–81). Springer International Publishing.
- Besson, S. (2011). Human rights and democracy in a global context: Decoupling and recoupling. *Ethics & Global Politics*, 4(1), 19–50. <https://doi.org/10.3402/egp.v4i1.6348>
- Bhat, P. (2021). Platform politics: The emergence of alternative social media in India. *Asia Pacific Media Educator*, 31(2), 269–276. <https://doi.org/10.1177/1326365X211056699>
- Bhat, P., & Chadha, K. (2020). Anti-media populism: Expressions of media distrust by right-wing media in India. *Journal of International and Intercultural Communication*, 13(2), 166–182. <https://doi.org/10.1080/17513057.2020.1739320>
- Bhatia, K. V., Elhussein, M., Kreimer, B., & Snapp, T. (2023). Protests, internet shutdowns, and disinformation in a transitioning state. *Media, Culture & Society*, 45(6), 1101–1118. <https://doi.org/10.1177/01634437231155568>
- Biddle, S. (2024). Elon Musk fought government surveillance – While profiting off government surveillance. *The Intercept*, 25 March. <https://theintercept.com/2024/03/25/elon-musk-x-dataminr-surveillance-privacy>
- Biju, P. R., & Gayathri, O. (2023). Self-breeding fake news: Bots and artificial intelligence perpetuate social polarization in India's conflict zones. *The International Journal of Information, Diversity, & Inclusion*, 7(1/2), 1–25. <https://doi.org/10.33137/ijidi.v7i1/2.39409>
- Bimber, B., & Gil de Zúñiga, H. (2020). The unedited public sphere. *New Media & Society*, 22(4), 700–715. <https://doi.org/10.1177/1461444819893980>
- Bineham, J. L. (1988). A historical account of the hypodermic model in mass communication. *Communication Monographs*, 55(3), 230–246. <https://doi.org/10.1080/03637758809376169>
- Birkinbine, B. J. (2018). Commons praxis: Towards a critical political economy of the digital commons. *tripleC*, 16(1), 290–305. <https://doi.org/10.31269/triplec.v16i1.929>
- Birnbaum, E., & Martin, E. (2023). Critics slam big tech lobbying in US Indo-Pacific talks. *TIME*, 10 July. <https://time.com/6293489/tech-firms-us-indo-pacific-trade-talks>

- Birwe, H. (2024). *La désinformation basée sur le genre en contexte de crises: Quels mécanismes de prévention et dispositifs de lutte pour les pays du Sahel?* IDOS Discussion Paper, Géographe diplômé de Sciences Po Lyon et de l'Université Cheikh Anta Diop de Dakar.
- Bitter, A. (2024). Amazon's just walk out: Technology relies on hundreds of workers in India watching you shop. *Business Insider India*, 3 August. www.businessinsider.in/retail/news/amazons-just-walk-out-technology-relies-on-hundreds-of-workers-in-india-watching-you-shop/articleshow/109014034.cms
- Blake, E., Odendaal, N., & Söderström, O. (2023). Beyond 'Data Positivism': Civil Society Organizations' Data and Knowledge Tactics in South Africa. In O. Söderström & A. Datta (Eds), *Data Power in Action* (pp. 204–226). Bristol University Press.
- Blasi Casagran, C., & Vermeulen, M. (2021). Reflections on the murky legal practices of political micro-targeting from a GDPR perspective. *International Data Privacy Law*, 11(4), 348–359. <https://doi.org/10.1093/idpl/ipab018>
- Blöbaum, B. (Ed.) (2016). *Trust and Communication in a Digitized World: Models and Concepts of Trust Research*. Springer International Publishing.
- Boadle, A. (2023). Brazil pushes back on big tech firms' campaign against 'fake news law'. Reuters, 3 May. www.reuters.com/world/americas/brazil-lawmakers-vote-controversial-bill-clean-up-social-media-2023-05-02
- Boas, T. C. (2006). Weaving the Authoritarian Web: The Control of Internet Use in Nondemocratic Regimes. In J. Zysman & A. Newman (Eds), *How Revolutionary Was the Digital Revolution? National Responses, Market Transitions and Global Technology* (pp. 361–378). Stanford University Press.
- Boczkowski, P. J., & Mitchelstein, E. (2021). *The Digital Environment: How We Live, Learn, Work, and Play Now*. MIT Press.
- Boczkowski, P. J., & Mitchelstein, E. (2022). *El Entorno Digital: Breve manual para entender cómo vivimos, aprendemos, trabajamos y pasamos el tiempo libre hoy*. Siglo Veintiuno.
- Boix, C. (2022). AI and the Economic and Informational Foundations of Democracy. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, et al. (Eds), *The Oxford Handbook of AI Governance* (Chapter 35). Oxford University Press.
- Bolin, G., & Kunelius, R. (2023). The return of propaganda: Historical legacies and contemporary conceptualisations. *Nordic Journal of Media Studies*, 5(1), 1–16. <https://doi.org/10.2478/njms-2023-0001>
- Bonfanti, M. E. (2020). *The Weaponisation of Synthetic Media: What Threat Does This Pose to National Security?* Center for Security Studies (CSS) at Zürich, Elcano Royal Institute for International and Strategic Studies (Real Instituto Elcano).
- Bonini, T., & Tréré, E. (2024). *Algorithms of Resistance: The Everyday Fight Against Platform Power*. MIT Press.
- Bontcheva, K., Papadopoulous, S., Tsalakanidou, F., Dutkiewicz, L., et al. (2024). *Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities*. Vera.ai, AI4Trust, AI4Media, & TITAN. <https://edmo.eu/wp-content/uploads/2023/12/Generative-AI-and-Disinformation-White-Paper-v8.pdf>
- Bontcheva, K., Posetti, J., Teyssou, D., Meyer, T., et al. (2020). *Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression: Broadband Commission Research Report on Freedom of Expression and Addressing Disinformation on the Internet*. Broadband Commission, ITU and UNESCO. www.broadbandcommission.org/publication/balancing-act-countering-digital-disinformation
- Bontridder, N., & Pouillet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy*, 3(e32), 1–21. <https://doi.org/10.1017/dap.2021.20>
- Boots, B. C., Matlack, A. K., & Richardson-Gool, T. S. (2024). A call for promoting algorithmic literacy. *Proceedings of the TPRC2024, the Research Conference on Communications, Information and Internet Policy*. <https://doi.org/https://papers.ssrn.com/abstract=4912427>
- Borelli, M. (2023). Social media corporations as actors of counter-terrorism. *New Media & Society*, 25(11), 2877–2897. <https://doi.org/10.1177/14614448211035121>
- Borges, D., & Reviglio, U. (2024). Diversification is good: Emerging business models for local news media. EUI (Centre for Media Pluralism and Media Freedom, 27 March. <https://cmpf.eui.eu/emerging-business-models-for-local-media>
- Borgogno, O., & Zangrandi, M. S. (2024). Chinese data governance and trade policy: From cyber sovereignty to the quest for digital hegemony? *Journal of Contemporary China*, 33(148), 578–602. www.tandfonline.com/doi/full/10.1080/10670564.2023.2299961
- Borocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press.
- Borz, G., De Francesco, F., Montgomerie, T. L., & Bellis, M. P. (2024). The EU soft regulation of digital campaigning: Regulatory effectiveness through platform compliance to the code of practice on disinformation. *Policy Studies, Online First*(0), 1–21. <https://doi.org/10.1080/01442872.2024.2302448>
- Bos, L., Kruijemeier, S., & de Vreese, C. (2016). Nation binding: How public service broadcasting mitigates political selective exposure. *PLOS ONE*, 11(5), 1–11. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155112>
- Botta, M. (2021). Sector regulation of digital platforms in Europe: Uno, Nessuno e Centomila. *Journal of European Competition Law & Practice*, 12(7), 500–512. <https://doi.org/10.1093/jeclap/lpab046>
- Boudry, C., Alvarez-Muñoz, P., Arencibia-Jorge, R., Ayena, D., Brouwer, N. J., Chaudhuri, Z., et al. (2019). *Worldwide inequality in access to full text scientific articles: The example of ophthalmology*. *PeerJ*, 7, 1–18. <https://doi.org/10.7717/peerj.7850>
- Boukes, M., & Vliegthart, R. (2017). News consumption and its unpleasant side effect: Studying the effect of hard and soft news exposure on mental well-being over time. *Journal of Media Psychology*, 29(3), 137–147. <https://doi.org/10.1027/1864-1105/a000224>

- Boulianne, S., Koc-Michalska, K., & Bimber, B. (2020). Right-wing populism, social media and echo chambers in Western democracies. *New Media & Society*, 22(4), 683–699. <https://doi.org/10.1177/1461444819893983>
- Boulianne, S., Tenove, C., & Buffie, J. (2022). Complicating the resilience model: A four-country study about misinformation. *Media and Communication*, 10(3), 169–182. <https://doi.org/10.17645/mac.v10i3.5346>
- Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, 10(1), 1–14. <https://doi.org/10.1038/s41467-018-07761-2>
- boyd, d. (2014). *It's Complicated: The Social Lives of Networked Teens*. Yale University Press.
- Bradford, A. (2020). *The Brussels Effect: How the European Union Rules the World*. Oxford University Press.
- Bradshaw, S., & DeNardis, L. (2022). Internet infrastructure as an emerging terrain of disinformation. CIGI (Centre for International Governance Innovation), Canada. www.cigionline.org/articles/internet-infrastructure-as-an-emerging-terrain-of-disinformation
- Bradshaw, S., & Howard, P. N. (2019). *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*. Oxford Internet Institute and University of Oxford. www.oii.ox.ac.uk/news-events/reports/the-global-disinformation-order-2019-global-inventory-of-organised-social-media-manipulation
- Bradshaw, S., Bailey, H., & Howard, P. N. (2021). *Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation*. Oxford Internet Institute, University of Oxford Project on Computational Propaganda.
- Bradshaw, S., Campbell-Smith, U., Henle, A., Perini, A., et al. (2020). *Country Case Studies, Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation*. Oxford Internet Institute, University of Oxford Project on Computational Propaganda. https://demotech.oii.ox.ac.uk/wp-content/uploads/sites/12/2021/03/Case-Studies_FINAL.pdf
- Brandtzaeg, P. B., Følstad, A., & Chaparro Domínguez, M. Á. (2018). How journalists and social media users perceive online fact-checking and verification services. *Journalism Practice*, 12(9), 1109–1129. <https://doi.org/10.1080/17512786.2017.1363657>
- Brayne, S. (2020). *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford University Press.
- Breuer, A. (2024). *Information Integrity and Information Pollution: Vulnerabilities and Impact on Social Cohesion and Democracy in Mexico*. IDOS Discussion Paper 2/2024. IDOS (German Institute of Development and Sustainability). www.idos-research.de/uploads/media/DP_2.2024.pdf
- Brevini, B. (2021). *Is AI Good for the Planet?*. Polity Press.
- Brewster, T. (2024). Musk's X fired 80% of engineers working on trust and safety, Australian Government says. *Forbes*, 10 January. www.forbes.com/sites/ssbrewster/2024/01/10/elon-musk-fired-80-per-cent-of-twitter-x-engineers-working-on-trust-and-safety
- Bria, F. (2018). *Barcelona City Council Digital Plan: Government Measure Concerning Ethical Management and Accountable Data: Barcelona Data Commons*. Open Digitalisation Programme, Barcelona City Council Office for Technology and Digital Innovation.
- Briant, E. L. (2021). *Propaganda Machine: The Hidden Story of Cambridge Analytica and the Digital Influence Industry*. Bloomsbury Publishing PLC.
- Briant, E. L. (2024). Hack Attacks: How Cyber Intimidation and Conspiracy Theories Drive the Spiral of 'Secrecy Hacking'. In J. Steel & J. Petley (Eds), *The Routledge Companion to Freedom of Expression and Censorship* (pp. 285–295). Routledge.
- Bridgman, A., Merkley, E., Loewen, P. J., Owen, T., et al. (2020). The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media. *HKS Misinformation Review*, 1(Special Issue), 1–18. <https://misinforeview.hks.harvard.edu/article/the-causes-and-consequences-of-covid-19-misperceptions-understanding-the-role-of-news-and-social-media>
- Bright, L. F., Logan, K., & Lim, H. S. (2022). Social media fatigue and privacy: An exploration of antecedents to consumers' concerns regarding the security of their personal information on social media platforms. *Journal of Interactive Advertising*, 22(2), 125–140. <https://doi.org/10.1080/15252019.2022.2051097>
- Brignull, H. (2023). *Deceptive Patterns: Exposing the Tricks Tech Companies Use to Control You*. Testimonium Ltd.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Broeders, D., Cristiano, F., & Kaminska, M. (2023). In search of digital sovereignty and strategic autonomy: Normative power Europe to the test of its geopolitical ambitions. *JCMS: Journal of Common Market Studies*, 61(5), 1261–1280. <https://doi.org/10.1111/jcms.13462>
- Brogi, E., & Sjøvaag, H. (2023). *Good Practices for Sustainable New Media Financing*. Committee of Experts on Increasing Resilience of Media (MSI-RES), Council of Europe.
- Brogi, E., Borges, D., Carlini, R., Nenadic, I., et al. (2023). *The European Media Freedom Act: Media Freedom, Freedom of Expression and Pluralism*. European Parliament's Committee on Civil Liberties, Justice and Home Affairs, PE 747.930.
- Broniatowski, D. A., Simons, J. R., Gu, J., Jamison, A. M., & Abrams, L. C. (2023). The efficacy of Facebook's vaccine misinformation policies and architecture during the COVID-19 pandemic. *Science Advances*, 9(37), 1–17. <https://doi.org/10.1126/sciadv.adh2132>
- Broomfield, H., & Reutter, L. (2022). In search of the citizen in the datafication of public administration. *Big Data & Society*, 9(1), 1–14. <https://doi.org/10.1177/20539517221089302>

- Broughton Micova, S. B., & Jacques, S. (2020). The functions of data in the competition between audiovisual media and video sharing platforms for advertising. *Journal of Information Policy*, 10, 514–548. <https://doi.org/10.5325/jinfopoli.10.2020.0514>
- Broughton Micova, S., & Schnurr, D. (2024). *Systemic Risk in Digital Services: Benchmarks for Evaluating the Management of Risks to Electoral Processes: Report*. CERRE (Centre on Regulation in Europe).
- Brown, C., & Levin, J. (2020). Prioritizing original news reporting on Facebook. Meta, Facebook, 30 June. <https://about.fb.com/news/2020/06/prioritizing-original-news-reporting-on-facebook>
- Brown, I., & Marsden, C. T. (2023). *Regulating Code: Good Governance and Better Regulation in the Information Age*. MIT Press.
- Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness*. Duke University Press.
- Bryant, M. (2021). Is Facebook leading us on a journey to the metaverse? *The Observer*, 26 September. www.theguardian.com/technology/2021/sep/26/is-facebook-leading-us-on-a-journey-to-the-metaverse
- Buarque, B., & Zavershinskaia, P. (2022). The Far-Right Politics of ‘Truth’: An Exploratory Analysis of the ‘Truths’ Produced by AfD Kompakt and Patriotic Alternative. In V. A. Bruno (Ed.), *Populism and Far-Right: Trends in Europe* (pp. 25–48). Educatt.
- Budak, C., Nyhan, B., Rothschild, D. M., Thorson, E., & Watts, D. J. (2024). Misunderstanding the harms of online misinformation. *Nature*, 630(8015), 45–53. <https://doi.org/10.1038/s41586-024-07417-w>
- Bullock, J. B., Chen, Y.-C., Himmelreich, J., Hudson, V. M., et al. (2022). *The Oxford Handbook of AI Governance*. Oxford University Press.
- Bunting, M. (2018). *Keeping Consumers Safe Online: Legislating for Platform Accountability for Online Content*. Communications Chambers.
- Buolamwini, J. (2023). *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*. Random House.
- Burgess, J., & Hurcombe, E. (2019). Digital journalism as symptom, response, and agent of change in the platformed media environment. *Digital Journalism*, 7(3), 359–367. <https://doi.org/10.1080/21670811.2018.1556313>
- Burgess, M. (2016). Google’s DeepMind trains AI to cut its energy bills by 40%. *Wired*, 28 July. www.wired.com/story/google-deepmind-data-centres-efficiency
- Burkart, N., & Huber, M. F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, 70, 245–317. <https://doi.org/10.1613/jair.112228>
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12. <https://journals.sagepub.com/doi/epub/10.1177/2053951715622512>
- Business Standard. (2018). US tech firms to boost lobbying against India’s push to store data locally. 18 August. www.business-standard.com/article/international/us-tech-firms-to-boost-lobbying-against-india-s-push-to-store-data-locally-118081800348_1.html
- Buturoiu, R., Corbu, N., & Boğan, M. (2023). *Patterns of News Consumption in a High-Choice Media Environment*. Springer Nature Switzerland.
- Caballero, F. S., & Monje, D. I. (2024). Artificial Intelligence and Technological Accelerationism: A Critique of Cybernetic Ideology. In J. Pedro-Caranana, R. Gómez, T. F. Corrigan, & F. S. Caballero (Eds.), *Political Economy of Media and Communication: Methodological Approaches* (pp. 328–365). Routledge.
- CAIDP (Center for AI and Digital Policy). (2022). *Artificial Intelligence and Democratic Values Index*. www.caidp.org/reports/aidv-2022
- Calderaro, A., & Blumfelde, S. (2022). Artificial intelligence and EU security: The false promise of digital sovereignty. *European Security*, 31(3), 415–434. <https://doi.org/10.1080/09662839.2022.2101885>
- Calhoun, C., Gaonkar, D. P., & Taylor, C. (2022). *Degenerations of Democracy*. Harvard University Press.
- Calice, M. N., Bao, L., Freiling, I., Howell, E., et al. (2023). Polarized platforms? How partisanship shapes perceptions of ‘algorithmic news bias’. *New Media & Society*, 25(11), 2833–2854. <https://doi.org/10.1177/14614448211034159>
- Calo, R., & Citron, D. K. (2021). The automated administrative state: A crisis of legitimacy. *Emory Law Journal*, 70(4), 799–844. <https://scholarlycommons.law.emory.edu/elj/vol70/iss4/1>
- Camara, M. S., Banu, H., & Abeck, J. C. (2023). The African Youth and Social Media at the Crossroads of Information, Misinformation, and Disinformation. In K. Langmia (Ed.), *Black Communication in the Age of Disinformation: DeepFakes and Synthetic Media* (pp. 15–50). Springer International Publishing.
- Cammaerts, B. (2018). *The Circulation of Anti-Austerity Protest*. Palgrave Macmillan.
- Cammaerts, B. (2024). Defending democracy against populist neo-Fascist attacks: The role and problems of public sphere theory. *Javnost – The Public*, 31(1), 26–45. <https://doi.org/10.1080/13183222.2024.2310983>
- Cammaerts, B., & Mansell, R. (2020). Digital platform policy and regulation: Toward a radical democratic turn. *International Journal of Communication*, 14(2020), 135–154. <https://ijoc.org/index.php/ijoc/article/view/11182/2901>
- Campbell-Stephens, R. M. (2021). *Educational Leadership and the Global Majority: Decolonising Narratives*. Springer Nature International.
- Cano-Orón, L., Calvo, D., García, G. L., & Baviera, T. (2021). Disinformation in Facebook ads in the 2019 Spanish general election campaigns. *Media and Communication*, 9(1), 217–228. <https://doi.org/10.17645/mac.v9i1.3335>

- Caplan, R. (2023). Networked platform governance: The construction of the democratic platform. *International Journal of Communication*, 17(2023), 3541–3472. <https://ijoc.org/index.php/ijoc/article/view/20035>
- Carmi, E., & Yates, S. (2023). Data citizenship: Data literacies to challenge power imbalance between society and 'big tech'. *International Journal of Communication*, 17(2023), 3619–3637. <https://ijoc.org/index.php/ijoc/article/view/18823>
- Carroll, S. R., Rodriguez-Lonebear, D., & Martinez, A. (2019). Indigenous data governance: Strategies from the United States' Native nations. *Data Science Journal*, 18(31), 1–15. <https://doi.org/doi:10.5334/dsj-2019-031>
- Carson, A., & Gibbons, A. (2023). The big chill? How journalists and sources perceive and respond to fake news laws in Indonesia and Singapore. *Journalism Studies*, 24(14), 1819–1838. <https://doi.org/10.1080/1461670X.2023.2192299>
- Carson, A., & Wright, S. (2022). Fake news and democracy: Definitions, impact and response. *Australian Journal of Political Science*, 57(3), 221–230. <https://doi.org/10.1080/10361146.2022.2122778>
- Casati, R. (2013). *Contro Il Colonialismo Digitale*. Laterza.
- Casciani, D., & BBC Verify. (2024). Violent Southport protests reveal far-right organising tactics. BBC News, 2 August. www.bbc.com/news/articles/cl4y0453nv5o
- Casero-Ripollés, A., Tuñón, J., & Bouza-García, L. (2023). The European approach to online disinformation: Geopolitical and regulatory dissonance. *Humanities and Social Sciences Communications*, 10(1), 1–10. <https://doi.org/10.1057/s41599-023-02179-8>
- Castro, L., Strömbäck, J., Esser, F., Van Aelst, P., et al. (2022). Navigating high-choice European political information environments: A comparative analysis of news user profiles and political knowledge. *The International Journal of Press/Politics*, 27(4), 827–859. <https://doi.org/10.1177/19401612211012572>
- Catalán-Matamoros, D. (Ed.) (2023). *Disinformation and Fact-Checking in Contemporary Society*. ESIC.
- Caton, S., & Haas, C. (2020). Fairness in machine learning: A survey. ArXiv. <https://doi.org/10.48550/arXiv.2010.04053>
- Caulfield, M., Bayar, M. C., & Aske, A. B. (2023). *The 'New Elites' of X: Identifying the Most Influential Accounts Engaged in Hamas/Israel Discourse*. Center for an Informed Public, University of Washington.
- Center for Countering Digital Hate. (2021). *The Toxic Ten. How Ten Fringe Publishers Fuel 69% of Digital Climate Change Denial*. <https://counterhate.com/research/the-toxic-ten>
- Centre on Media Pluralism, Media Freedom, CITI, KU Leuven, et al. (2022). *Study on Media Pluralism and Diversity Online: Final Report*. European Commission, LC-01637953.
- CERTH (Center for Research and Technology Hellas), iti, The Left in the European Parliament. (2021). *Online Disinformation in Europe: A Study on Ethical Standards, Involved Political Interests and Policy Recommendations*. The Left in the European Parliament. <https://left.eu/issues/study-online-disinformation-in-europe>
- Chadha, K., & Bhat, P. (2022). Alternative news media and critique of mainstream journalism in India: The case of *OpIndia*. *Digital Journalism*, 10(8), 1283–1301. <https://doi.org/10.1080/21670811.2022.2118143>
- Chadwick, A. (2017). *The Hybrid Media System: Politics and Power*. Second Edition. Oxford University Press.
- Chaguaceda, A., Cilano Pelaez, J., & Puerta, M. I. (2023). Illiberal narratives in Latin America: Russian and allied media as vehicles of autocratic cooperation. *The Journal of Illiberalism Studies*, 3(2), 111–123. <https://doi.org/10.53483/XCMW3558>
- Chair, C. (2024). Gendered Digital Inequalities: How Do We Ensure Gender Transformative Law and Practice in the Age of Artificial Intelligence in Africa? In L. A. Abdulrauf & H. Dube (Eds), *Data Privacy Law in Africa: Emerging Perspectives* (pp. 207–238). Pretoria University Law Press.
- Chairpersons Organizations. (2023). *Statement of the Networks of Media and Online Safety Regulators on the Occasion of the UNESCO Global Conference Internet for Trust*. ACRAN (African Communication Regulation Authorities Network), ERGA (European Regulators Group for Audiovisual Media Services), GOSRN (Global Online Safety Regulators Network), MNRA (Mediterranean Network of Regulatory Authorities), PRAI (Platform of Ibero-American Audiovisual Regulators) and REFRAM (Francophone Network of Media Regulators) Statement by Chairs at UNESCO Global Conference – Internet for Trust. www.rirm.org/wp-content/uploads/2017/06/Statement-UNESCO-Global-Conference-en.pdf
- Chambers, R. (1997). *Whose Reality Counts? Putting the First Last*. Practical Action Publishing.
- Chambers, S. (2023). Deliberative democracy and the digital public sphere: Asymmetrical fragmentation as a political not a technological problem. *Constellations*, 30(1), 61–68. <https://doi.org/10.1111/1467-8675.12662>
- Chasi, C., & Rodny-Gumede, Y. (2022). Decolonial journalism: New notes on ubuntu and the public interest. *Journalism Studies*, 23(13), 1625–1637. <https://doi.org/10.1080/1461670X.2022.2083006>
- Chavalarias, D. (2024). Minuit moins dix à l'horloge de Poutine: Analyse de réseaux des ingérences étrangères dans les élections législatives de 2024. Hal Open Science. <https://hal.science/hal-04629585v3/document>
- Chen, J. Q., & Benusa, A. (2017). HIPAA security compliance challenges: The case for small healthcare providers. *International Journal of Healthcare Management*, 10(2), 135–146. <https://doi.org/10.1080/20479700.2016.1270875>
- Chen, Y., & Gao, Y. (2022). Comparative analysis of digital trade development strategies and governance approaches. *Journal of Digital Economy*, 1(3), 227–238. <https://doi.org/10.1016/j.jdec.2023.02.001>
- Cheng, D. (2016). *Cyber Dragon: Inside China's Information Warfare and Cyber Operations*. Bloomsbury Publishing USA.
- Chi, N., Lurie, E., & Mulligan, D. K. (2021). Reconfiguring diversity and inclusion for AI ethics. In *AIES '21: AAAI/ACM Conference on AI, Ethics, and Society* (pp. 447–457). <https://doi.org/10.1145/3461702.3462622>

- Chibwe, A. (2020). Social media and elections in Zimbabwe: Twitter war between pro-ZANU-PF and pro-MDC-A netizens. *South African Journal of Communication Theory and Research*, 46(4), 7–30. <https://doi.org/10.1080/02500167.2020.1723663>
- Chin, Y. C. (2024). Right to Data Access in the Digital Era: The Case of China. In M. Aslama Horowitz, H. Niemeinen, K. Lehtisaari, & A. D'Arma (Eds), *Epistemic Rights in the Era of Digital Disruption* (pp. 187–201). Springer International Publishing.
- Chiumbu, S., & Munoriyarwa, A. (2023). Exploring data journalism practices in Africa: Data politics, media ecosystems and newsroom infrastructures. *Media, Culture & Society*, 45(4), 841–858. <https://doi.org/10.1177/01634437231155341>
- Cho, H., Cannon, J., Lopez, R., & Li, W. (2024). Social media literacy: A conceptual framework. *New Media & Society*, 26(2), 941–960. <https://doi.org/10.1177/14614448211068530>
- Choi, J., Kim, K. K., & Kim, Y. (2020). Emotional Analysis with News Using Text Mining for Framing Theory. In R. Lee (Ed.), *Computer and Information Science* (pp. 95–108). Springer International Publishing.
- Chonka, P., Diepeveen, S., & Haile, Y. (2023). Algorithmic power and African indigenous languages: Search engine autocomplete and the global multilingual Internet. *Media, Culture & Society*, 45(2), 246–265. <https://doi.org/10.1177/01634437221104705>
- Chouliarakis, L., & Georgiou, M. (2022). *The Digital Border: Migration, Technology, Power*. NYU Press.
- Christodoulou, E., & Iordanou, K. (2021). Democracy under attack: Challenges of addressing ethical issues of AI and big data for more democratic digital media and societies. *Frontiers in Political Science*, 3, 1–17. www.frontiersin.org/articles/10.3389/fpos.2021.682945/full
- Christofoletti, R., & Becker, D. (2023). Pesquisas sobre confiança na mídia e o fortalecimento da ideia de crise permanente de credibilidade. *Chasqui: Revista Latinoamericana de Comunicación*, 154, 53–67. <https://dialnet.unirioja.es/servlet/articulo?codigo=9274150>
- Christopher, N. (2021). Amazon's 'Delivery Premier League' gamifies gig work in India. Rest of World Reporting Global Tech Stories, 25 October. <https://restofworld.org/2021/amazons-delivery-premier-league-gamifies-gig-work-in-india>
- Chung, M., & Wihbey, J. (2024). Social media regulation, third-person effect, and public views: A comparative study of the United States, the United Kingdom, South Korea, and Mexico. *New Media & Society*, 26(8), 4534–4553. <https://doi.org/10.1177/14614448221122996>
- Church, P., & Pehlivan, C. N. (2023). The Digital Services Act (DSA): A new era for online harms and intermediary liability. *Global Privacy Law Review*, 4(1), 1–7. <https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\GPLR\GPLR2023005.pdf>
- Cieslik, K., & Margócsy, D. (2022). Datafication, power and control in development: A historical perspective on the perils and longevity of data. *Progress in Development Studies*, 22(4), 352–373. <https://doi.org/10.1177/14649934221076580>
- Cipers, S., Meyer, T., & Lefevere, J. (2023). Government responses to online disinformation unpacked. *Internet Policy Review*, 12(4). <https://policyreview.info/articles/analysis/government-responses-to-online-disinformation-unpacked>
- Claesson, A. (2023). Twitter: A necessary evil? Journalistic responses to Elon Musk and the denormalization of social media. *Journalism, Online First*, 1–18. <https://doi.org/10.1177/14648849231221616>
- CNTI (Center for News, Technology & Innovation). (2024). Most 'Fake News' Legislation Risks Doing More Harm than Good amid a Record Number of Elections in 2024. 3 September. <https://innovating.news/article/most-fake-news-legislation-risks-doing-more-harm-than-good-amid-a-record-number-of-elections-in-2024>
- Cohen, J. E. (2019). *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford University Press.
- Cohen, J. E. (2023). Infrastructuring the digital public sphere. *Yale Journal of Law and Technology*, 25(1), 1–40. https://law.yale.edu/sites/default/files/area/center/isp/documents/cohen_julie_-_infrastructuring_the_digital_public_sphere.01.pdf
- Cohn, A., Stoop, J., & Rahman, H. A. (2022). *Disinformation for Hire: Examining the Production of False COVID-19 Information*. Tinbergen Institute Discussion Paper, TI 2022-086/II. <https://papers.tinbergen.nl/22086.pdf>
- Colclough, C. (2022). Reshaping the digitization of public services. *New England Journal of Public Policy*, 34(1), 1–14. <https://scholarworks.umb.edu/nejpp/vol34/iss1/9>
- Cole, M. D. (2024). AI regulation and governance on a global scale: An overview of international, regional and national instruments. *Journal of AI Law and Regulation*, 1(1), 126–142. <https://doi.org/10.21552/aire/2024/1/16>
- Collao, K. (2024). *Ok Computer? Public Attitudes to the Uses of Generative AI in News*. Craft and Reuters Institute for the Study of Journalism, Oxford University.
- Conroy-Krutz, J., & Koné, J. (2022). *Confronting Threats to Civic Spaces: Final Report*. Afrobarometer. www.afrobarometer.org/wp-content/uploads/2023/05/NED-Confronting-Threats-Report-Afrobarometer-Final-27june22.pdf
- Cooke, D. (2023). *Synthetic Media and Election Integrity: Defending Our Democracies*. Alan Turing Institute, Centre for Emerging Technology and Security (CETaS).
- Cooke, P. (2020). Silicon Valley imperialists create new model villages as smart cities in their own image. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(24), 1–18. <https://doi.org/10.3390/joitmc6020024>
- Coombs, W. T., & Holladay, S. J. (Eds) (2022). *The Handbook of Crisis Communication*, Second Edition. John Wiley & Sons.

- Coos, A. (2022). Data protection in Japan: All you need to know about APPI. Endpoint Protector Blog, 5 April. www.endpointprotector.com/blog/data-protection-in-japan-appi
- Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. *ArXiv*. <https://doi.org/10.48550/ArXiv.1701.08230>
- Corbu, N., Bărgăoanu, A., Durach, F., & Udrea, G. (2021). Fake news going viral: The mediating effect of negative emotions. *Media Literacy and Academic Research*, 4(2), 58–87. www.mlar.sk/wp-content/uploads/2021/12/4_Corbu.pdf
- Cornils, M. (2020). *Designing Platform Governance: A Normative Perspective on Needs, Strategies, and Tools to Regulate Intermediaries*. AlgorithmWatch, European Policy Centre, Mainzer Medieninstitut and Civitates.
- Corporate Europe Observatory. (2024a). Don't let corporate lobbying further water down the AI Act, lobby watchdogs warn MEPs. 12 March. <https://corporateeurope.org/en/2024/03/dont-let-corporate-lobbying-further-water-down-ai-act-lobby-watchdogs-warn-meps>
- Corporate Europe Observatory. (2024b). Trojan horses: How European startups teamed up with Big Tech to gut the AI Act. 11 March. [www.corporateeurope.org/en/2024/03/trojan-horses-how-european-startups-teamed-big-tech-gut-ai-act](https://corporateeurope.org/en/2024/03/trojan-horses-how-european-startups-teamed-big-tech-gut-ai-act)
- Corrales Compagnucci, M., Wilson, M. L., Fenwick, M., Forgó, N., & Bärnighausen, T. (Eds) (2022). *AI in eHealth: Human Autonomy, Data Governance and Privacy in Healthcare*. Cambridge University Press.
- Couldry, N., & Curran, J. (Eds) (2003). *Contesting Media Power: Alternative Media in a Networked World*. Rowman & Littlefield.
- Couldry, N., & Hepp, A. (2016). *The Mediated Construction of Reality: Society, Culture, Mediatization*. Polity.
- Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media*, 20(4), 336–349. <https://doi.org/10.1177/1527476418796632>
- Council of Europe. (2018). *Recommendation CM/Rec(2018)1 of the Committee of Ministers to Member States on Media Pluralism and Transparency of Media Ownership*. CM/Rec(2018)1.
- Council of Europe. (2022). *Recommendation CM/Rec(2022)12 of the Committee of Ministers to Member States on Electoral Communication and Media Coverage of Election Campaigns*. CM/Rec(2022)12.
- Council of Europe. (2023). *Guidelines for the Responsible Implementation of Artificial Intelligence (AI) in Journalism*, <https://www.coe.int/en/web/freedom-expression/-/guidelines-on-the-responsible-implementation-of-artificial-intelligence-ai-systems-in-journalism>
- Council of Europe. (2024). *Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*. 133rd Meeting, 17 May 2024, CM(2024)52-final.
- Cows, J., Tsamados, A., Taddeo, M., & Floridi, L. (2023). The AI gambit: Leveraging artificial intelligence to combat climate change – Opportunities, challenges, and recommendations. *AI & SOCIETY*, 38(1), 283–307. <https://doi.org/10.1007/s00146-021-01294-x>
- Craig, J., Simmons, C., & Bhatnagar, R. (2023). How January 6 inspired election disinformation around the world. *ISD Digital Dispatches*, 13 January. www.isdglobal.org/digital_dispatches/how-january-6-inspired-election-disinformation-around-the-world
- Cramer, B. W. (2020). From liability to accountability: The ethics of citing Section 230 to avoid the obligations of running a social media platform. *Journal of Information Policy*, 10, 123–150. <https://doi.org/10.5325/jinfopoli.10.2020.0123>
- Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Crawford, K., Gray, M. L., & Miltner, K. (2014). Critiquing big data: Politics, ethics, epistemology. *International Journal of Communication*, 8(2014), 1663–1672. <https://ijoc.org/index.php/ijoc/article/view/2167>
- Crémer, J., de Montjoye, Y.-A., & Schweitzer, H. (2019). *Competition Policy for the Digital Era: Final Report*. European Commission DG Competition.
- Crosset, V., & Dupont, B. (2022). Cognitive assemblages: The entangled nature of algorithmic content moderation. *Big Data & Society*, 9(2), 1–13. <https://doi.org/10.1177/20539517221143361>
- Crump, C. (2016). Surveillance policy making by procurement. *Washington Law Review*, 91(4), 1595–1662. <https://doi.org/10.2139/ssrn.2737006>
- Culloty, E. (2021). Désinformation sur la migration: UN problème de longue date au dimensions technologiques nouvelles. In M. McAuliffe & A. Triandafyllidou (Eds), *Rapport État de la Migration Dans Le Monde* (pp. 229–245). Organisation internationale pour les migrations (OIM).
- Cunliffe-Jones, P. (2021). How government responses to misinformation in Africa restrict freedom of expression and do little to tackle the problem. *African Journalism Studies*, 42(2), 121–125. <https://doi.org/10.1080/23743670.2021.1972532>
- Cunliffe-Jones, P. (2021). How government responses to misinformation in Africa restrict freedom of expression and do little to tackle the problem. *African Journalism Studies*, 42(2), 121–125. <https://doi.org/10.1080/23743670.2021.1972532>
- Curran, J., & Hesmondhalgh, D. (Eds) (2019). *Media and Society*. 6th Edition. Bloomsbury.
- Curran, J., & Park, M.-J. (Eds) (1999). *De-Westernizing Media Studies*. Routledge.
- Curtis, J. S. (2021). Springing the 'Tacitus Trap': Countering Chinese state-sponsored disinformation. *Small Wars & Insurgencies*, 32(2), 229–265. <https://doi.org/10.1080/09592318.2021.1870429>

- Curzi de Mendonça, Y. (2024). Elon Musk's feud with Brazilian judge is much more than a personal spat – it's about national sovereignty, freedom of speech and the rule of law. *The Conversation*, 9 September. <https://theconversation.com/elon-musks-feud-with-brazilian-judge-is-much-more-than-a-personal-spat-its-about-national-sovereignty-freedom-of-speech-and-the-rule-of-law-238264>
- Cushion, S. (2021). Are public service media distinctive from the market? Interpreting the political information environments of BBC and commercial news in the United Kingdom. *European Journal of Communication*, 37(1), 3–20. <https://doi.org/10.1177/02673231211012149>
- Custers, B., & Fosch-Villaronga, E. (Eds) (2022). *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*. T.M.C. Asser Press.
- Cyphers, B. (2022). Inside Fog Data Science, the secretive company selling mass surveillance to local police. Electronic Frontier Foundation, 31 August. www.eff.org/deeplinks/2022/08/inside-fog-data-science-secretive-company-selling-mass-surveillance-local-police
- d'Haenens, L., Lo, W. H., & Moore, M. (2022). Innovation in Journalism: How Technology Affects the News Media, Publication Formats, and the Journalist Profession. In J. Trappel & T. Tomaz (Eds), *Success and Failure in News Media Performance: Comparative Analysis in the Media for Democracy Monitor 2021* (pp. 337–345). Nordicom.
- da Silva, C. R. S., Teixeira, T. M. C., & Pinto, V. B. (2019). Research methodology in information literacy: A systematic review. *RDBCI: Revista Digital de Biblioteconomia e Ciência Da Informação*, 17, 1–17. <https://doi.org/10.20396/rdbci.v17i0.8653728/eO19014>
- Dadakhonov, A. O. (2024). Analysis of media and information literacy definitions: A qualitative approach. *Studies in Media and Communications*, 12(2), 116–129. <https://redfame.com/journal/index.php/smc/article/view/6659>
- Dahlberg, L. (2014). The Habermasian public sphere and exclusion: An engagement with poststructuralist-influenced critics. *Communication Theory*, 24(1), 21–41. <https://doi.org/10.1111/comt.12010>
- DAIR (Distributed Artificial Intelligence Research) Institute. (no date). Our research is intended to benefit communities. www.dair-institute.org/research-philosophy
- Dalle, J.-M., & David, P. A. (2007). Simulating Code Growth in Libre (Open Source) Mode. In E. Brousseau & N. Curien (Eds), *Internet and Digital Economics: Principles, Methods and Applications* (pp. 391–422). Cambridge University Press.
- Daniels, J. (2018). The algorithmic rise of the 'alt-right'. *Contexts*, 17(1), 60–65. <https://doi.org/10.1177/1536504218766547>
- Das, A., & Schroeder, R. (2021). Online disinformation in the run-up to the Indian 2019 election. *Information, Communication & Society*, 24(12), 1762–1778. <https://doi.org/10.1080/1369118X.2020.1736123>
- Davis, A., Fenton, N., Freedman, D., & Khiabany, G. (2020). *Media, Democracy and Social Change: Re-imagining Political Communications*. SAGE Publications.
- De Abreu, B. S. (Ed.) (2022). *Media Literacy, Equity, and Justice*. Taylor & Francis.
- De Blasio, E., & Selva, D. (2021). Who is responsible for disinformation? European approaches to social platforms' accountability in the post-truth era. *American Behavioral Scientist*, 65(6), 825–846. <https://doi.org/10.1177/0002764221989784>
- De Bruyne, J., & Vanleenhove, C. (Eds) (2021). *Artificial Intelligence and the Law*. Intersentia.
- De Gregorio, G. (2023). The normative power of artificial intelligence. *Indiana Journal of Global Legal Studies*, 30(2), 55–80. <https://ssrn.com/abstract=4436287>
- De Gregorio, G., & Dunn, P. (2023). Artificial Intelligence and Freedom of Expression. In A. Quintavalla & J. Temperman (Eds), *Artificial Intelligence and Human Rights* (pp. 76–90). Oxford University Press.
- De Gregorio, G., & Stremlau, N. (2021). Platform Governance at the Periphery: Moderation, Shutdowns and Intervention. In J. Bayer, B. Holznagel, P. Korpisaari, & L. Woods (Eds), *Perspectives on Platform Regulation: Concepts and Models of Social Media Governance Across the Globe* (pp. 433–450). Nomos Verlagsgesellschaft mbH & Co. KG.
- De Gregorio, G., & Stremlau, N. (2023). Inequalities and content moderation. *Global Policy*, 14(5), 870–879. <https://doi.org/10.1111/1758-5899.13243>
- de-Lima Santos, M.-F. (2023). The entanglements between data journalism, collaboration and business models: A systematic literature review. *Digital Journalism*, 12(2), 256–281. <https://doi.org/10.1080/21670811.2023.2247449>
- de Lima Santos, M.-F., Yeung, W. N., & Dodds, T. (2024). Guiding the way: A comprehensive examination of AI guidelines in global media. *AI & SOCIETY*, 1–19. <https://doi.org/10.1007/s00146-024-01973-5>
- De Paor, S., & Heravi, B. (2020). Information literacy and fake news: How the field of librarianship can help combat the epidemic of fake news. *The Journal of Academic Librarianship*, 46(5), 1–8. <https://doi.org/10.1016/j.acalib.2020.102218>
- de Sousa Santos, B. (2018). *The End of the Cognitive Empire: The Coming of Age of Epistemologies of the South*. Duke University Press.
- de Vivo, I. (2023). The 'neo-intermediation' of large on-line platforms: Perspectives of analysis of the 'state of health' of the digital information ecosystem. *Communications*, 48(3), 420–439. <https://doi.org/10.1515/commun-2022-0102>
- Dellinger, J., & Pell, S. (2024). Bodies of evidence: The criminalization of abortion and surveillance of women in a post-DOBBS world. *Duke Journal of Constitutional Law & Public Policy*, 19, 1–99. <http://dx.doi.org/10.2139/ssrn.4599445>
- Demeter, M. (2019). The winner takes it all: International inequality in communication and media studies today. *Journalism & Mass Communication Quarterly*, 96(1), 37–59. <https://doi.org/10.1177/1077699018792270>

- Dencik, L., & Leistert, O. (Eds) (2015). *Critical Perspectives on Social Media and Protest: Between Control and Emancipation*. Rowman & Littlefield.
- Dencik, L., Hintz, A., & Cable, J. (2016). Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society*, Jul-Dec, 1-12. <https://journals.sagepub.com/doi/pdf/10.1177/2053951716679678>
- Dergacheva, D., Kuznetsova, V., Scharlach, R., & Katzenbach, C. (2023). *One Day in Content Moderation: Analyzing 24 h of Social Media Platforms' Content Decisions through the DSA Transparency Database*. University of Bremen.
- Deuze, M. (2014). *Media Life*. Polity Press.
- Deuze, M., & Beckett, C. (2022). Imagination, algorithms and news: Developing AI literacy for journalism. *Digital Journalism*, 10(10), 1913-1918. <https://doi.org/10.1080/21670811.2022.2119152>
- Devenney, M. (2009). The limits of communicative rationality and deliberative democracy. *Journal of Power*, 2(1), 137-154. <https://doi.org/10.1080/17540290902760915>
- Dias Oliva, T. (2020). Content moderation technologies: Applying human rights standards to protect freedom of Expression. *Human Rights Law Review*, 20(4), 607-640. <https://doi.org/10.1093/hrlr/ngaa032>
- Diaz Ruiz, C. (2023). Disinformation on digital media platforms: A market-shaping approach. *New Media & Society*, Online First, 1-24. <https://doi.org/10.1177/14614448231207644>
- Díaz, Á., & Hecht-Felella, L. (2021). *Double Standards in Social Media Content Moderation*. Brennan Center for Justice, New York University School of Law.
- Díaz, A., & Levinson-Waldman, R. (2020). *Automatic License Plate Readers: Legal Status and Policy Recommendations for Law Enforcement Use*. Brennan Center for Justice. www.brennancenter.org/our-work/research-reports/automatic-license-plate-readers-legal-status-and-policy-recommendations
- Dibenedetto, C. (2021). Faced with limited resources, Indigenous communities built their own internet. Here's how. Indigenous Peoples Major Group for Sustainable Development. <https://indigenouspeoples-sdg.org/index.php/english/ttt/1549-faced-with-limited-resources-indigenous-communities-built-their-own-internet-here-s-how>
- Dierickx, L., Lindén, C.-G., & Opdahl, A. L. (2023c). The Information Disorder Level (IDL) index: A Human-Based Metric to Assess the Factuality of Machine-Generated Content. In D. Ceolin, T. Caselli, & M. Tulin (Eds), *Disinformation in Open Online Media* (pp. 60-71). Springer-Verlag.
- Dierickx, L., Lindén, C.-G., & Opdahl, A. L. (2023a). Automated fact-checking to support professional practices: Systematic literature review and meta-analysis. *International Journal of Communication*, 17(2023), 5170-5190. <https://ijoc.org/index.php/ijoc/article/view/21071>
- Dierickx, L., Linden, C.-G., Opdahl, A., Khan, S. A., & Rojas, D. (2023b). AI in the newsroom: A data quality assessment framework for employing machine learning in journalistic workflows. *Proceedings of the 5th International Conference on Advanced Research Methods and Analytics*. <http://hdl.handle.net/10251/201692>
- DiGi (Digital Industry Group). (2022). *Australian Code of Practice on Disinformation and Misinformation*. <https://digi.org.au/disinformation-code>
- DiResta, R. (2024). My encounter with the fantasy-industrial complex. *The Atlantic*, 15 June. www.theatlantic.com/ideas/archive/2024/06/cia-renee-censorship-conspiracy-twitter/678688
- Dittrich, P.-J. (2019). *Tackling the Spread of Disinformation: Why a Co-Regulatory Approach Is the Right Way Forward for the EU*. Bertelsmann Stiftung Policy Paper 12, Hertie School, Jacques Delors Centre.
- Doctorow, C., & Giblin, R. (2023). *Chokepoint Capitalism: How Big Tech and Big Content Captured Creative Labor Markets and How We'll Win Them Back*. Beacon Press.
- Dolata, M., Feuerriegel, S., & Schwabe, G. (2022). A sociotechnical view of algorithmic fairness. *Information Systems Journal*, 32(4), 754-818. <https://doi.org/10.1111/isj.12370>
- Domingos Cordeiro, V., & Cozman, F. (2024). Artificial Intelligence and Everyday Knowledge. In H. S. Dunn, M. Ragnedda, M. L. Riuu, & L. Robinson (Eds), *The Palgrave Handbook of Everyday Digital Life* (pp. 23-35). Springer Link.
- Dommett, K. (2023). The inter-institutional impact of digital platform companies on democracy: A case study of the UK media's digital campaigning coverage. *New Media & Society*, 25(10), 2763-2780. <https://doi.org/10.1177/14614448211028546>
- Douai, A. (2019). Global and Arab Media in the Post-Truth Era: Globalization, Authoritarianism and Fake News. In *IEMed: Mediterranean Yearbook 2019* (pp. 124-132). IEMed – European Institute of the Mediterranean.
- Douek, E. (2024). The Meta Oversight Board and the empty promise of legitimacy. *Harvard Journal of Law & Technology*, 37(2), 373-445. <http://dx.doi.org/10.2139/ssrn.4565180>
- Dowling, M.-E. (2022). Cyber information operations: Cambridge Analytica's challenge to democratic legitimacy. *Journal of Cyber Policy*, 7(2), 230-248. <https://doi.org/10.1080/23738871.2022.2081089>
- DPGA (Digital Public Goods Alliance). (2023). *Collaborative Momentum: The 2023 State of the Digital Public Goods Ecosystem Report*.
- Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729-745. <https://doi.org/10.1080/1369118X.2018.1428656>
- Duffy, A., & Ang, P. H. (2019). Digital journalism: defined, refined, or re-defined. *Digital Journalism*, 7(3), 378-385. <https://doi.org/10.1080/21670811.2019.1568899>

- Dugan, L., Ippolito, D., Kirubakaran, A., Shi, S., & Callison-Burch, C. (2023). Real or fake text? Investigating human ability to detect boundaries between human-written and machine-generated text. *ArXiv*. <https://doi.org/10.48550/arXiv.2212.12672>
- Duncan, J. (2023). Data protection beyond data rights: Governing data production through collective intermediaries. *Internet Policy Review*, 12(3), 1–22. <https://doi.org/10.14763/2023.3.1722>
- Dunn, H. S., Ragnedda, M., Ruiu, M. L., & Robinson, L. (Eds) (2024). *The Palgrave Handbook of Everyday Digital Life*. Palgrave Macmillan.
- Durach, F., Bărgăoanu, A., & Nastasiu, C. (2020). Tackling disinformation: EU regulation of the digital space. *Romanian Journal of European Affairs*, 20(1), 5–20. <https://papers.ssrn.com/abstract=3650780>
- Durach, F., Vladu, L., Călin, I., & Enache, M. (2024). Public Policies to Fight Disinformation Actively through Media and Information Literacy. In D. Frau-Meigs & N. Corbu (Eds), *Disinformation Debunked* (Chapter 10). Routledge.
- Dutta, M. J., & Pal, M. (2020). Theorizing from the Global South: Dismantling, resisting, and transforming communication theory. *Communication Theory*, 30(4), 349–369. <https://doi.org/10.1093/ct/qtaa010>
- Dutta, S., & Lanvin, B. (2023). *Network Readiness Index 2023: Trust in a Network Society: A Crisis of the Digital Age?* Portulans Institute, University of Oxford.
- Dutton, W. H. (2023). *The Fifth Estate: The Power Shift of the Digital Age*. Oxford University Press.
- Dyson, I. (2023). New York City must strengthen police transparency law. Brennan Center for Justice, 15 December. www.brennancenter.org/our-work/analysis-opinion/new-york-city-must-strengthen-police-transparency-law
- Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. (2019). How many people live in political bubbles on social media? Evidence From linked survey and Twitter data. *SAGE Open*, 9(1), 1–21. <https://doi.org/10.1177/2158244019832705>
- Eady, G., Paskhalis, T., Zilinsky, J., Bonneau, R., Nagler, J., & Tucker, J. A. (2023). Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. *Nature Communications*, 14(1), 1–11. <https://doi.org/10.1038/s41467-022-35576-9>
- Eastwood, B. (2024). Study: Industry now dominates AI research. MIT Management Sloan School, 18 May. <https://mitsloan.mit.edu/ideas-made-to-matter/study-industry-now-dominates-ai-research>
- EC (European Commission). (2012). *Consolidated Version of the Treaty on the Functioning of the European Union (TFEU)*. C326/47. www.legislation.gov.uk/eut/teec/contents
- EC (2016a). *Background Note on Open Access to Scientific Publications and Open Research Data*. https://research-and-innovation.ec.europa.eu/document/download/4bd9ef8e-0101-457d-9fc5-1096c4e8f6f0_en?filename=ec_rtd_background-note-open-access.pdf
- EC (2016b). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)*. OJ L 119/1. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679>
- EC (2017). *Mapping of Media Literacy Practices and Actions in EU-28*. European Commission Audio-Visual Observatory.
- EC. (2018). *Communication – Tackling Online Disinformation: A European Approach*. COM(2018) 236 final. <https://digital-strategy.ec.europa.eu/en/library/communication-tackling-online-disinformation-european-approach>
- EC HLG (High Level Group). (2018). *A Multi-Dimensional Approach to Disinformation: Report of the Independent High Level Group on Fake News and Online Disinformation*. <https://op.europa.eu/en/publication-detail/-/publication/6ef4df8b-4cea-11e8-be1d-01aa75ed71a1/language-en>
- EC (2020a). *Assessment of the Code of Practice on Disinformation – Achievements and Areas for Further Improvement*. <https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement>
- EC. (2020b). *Communication on the European Democracy Action Plan*. COM(2020) 790 final. www.europeansources.info/record/communication-on-the-european-democracy-action-plan
- EC. (2022a). *Contestable and Fair Markets in the Digital Sector and Amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)*, PE/17/2022/REV/1. ERegulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022. <https://eur-lex.europa.eu/eli/reg/2022/1925/oj>
- EC. (2022b). *European Commission Digital Strategy*. European Commission C(2022) 4388 final. https://commission.europa.eu/publications/european-commission-digital-strategy_en
- EC. (2022c). *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)*. <https://eur-lex.europa.eu/eli/reg/2022/2065/oj>
- EC. (2022d). *Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act)*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R0868>
- EC. (2022e). *The Strengthened Code of Practice on Disinformation 2022*. <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>
- EC: Directorate-General for Communications Networks, Content and Technology, Parcu, P., Brogi, E., Verza, S., Da Costa Leite Borges, D., et al. (2022) *Study on Media Plurality and Diversity Online – Final Report*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2759/529019>

- EC. (2023). *Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on Harmonised Rules on Fair Access to and Use of Data and Amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202302854
- EC. (2024a). Commission requests information from X on decreasing content moderation resources under the Digital Services Act. European Commission News, 8 May. <https://digital-strategy.ec.europa.eu/en/news/commission-requests-information-x-decreasing-content-moderation-resources-under-digital-services>
- EC. (2024b). *Regulation (EU) 2024/1083 Establishing a Common Framework for Media Service (European Media Freedom Act)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401083
- EC. (2024c). *Regulation (EU) 2024/1689 of the European Parliament and of the Council Of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689
- EC. (2024d). Funded projects in the fight against disinformation. https://commission.europa.eu/strategy-and-policy/coronavirus-response/fighting-disinformation/funded-projects-fight-against-disinformation_en
- Ecker, U., Roozenbeek, J., van der Linden, S., Tay, L. Q., et al. (2024a). Misinformation poses a bigger threat to democracy than you might think. *Nature*, 630(8015), 29–32. <https://doi.org/10.1038/d41586-024-01587-3>
- Ecker, U., Tay, L. Q., Roozenbeek, J., van der Linden, S., et al. (2024b). Why misinformation must not be ignored. *PsyArXiv*. <https://doi.org/10.31234/osf.io/8a6cj>
- Ecker, U., Tay, L. Q., Roozenbeek, J., van der Linden, S., et al. (2024). Why misinformation must not be ignored. *PsyArXiv*. <https://doi.org/10.31234/osf.io/8a6cj>
- Economides, N., & Hermalin, B. E. (2012). The economics of network neutrality. *The RAND Journal of Economics*, 43(4), 602–629. www.jstor.org/stable/41723347
- Eddy, K. (2024). Republicans, young adults now nearly as likely to trust info from social media as from national news outlets. *Pew Research Center*. <https://www.pewresearch.org/short-reads/2024/10/16/republicans-young-adults-now-nearly-as-likely-to-trust-info-from-social-media-as-from-national-news-outlets/>
- Edgerly, S. (2022). The head and heart of news avoidance: How attitudes about the news media relate to levels of news consumption. *Journalism*, 23(9), 1828–1845. <https://doi.org/10.1177/14648849211012922>
- Effoduh, J. O. (2024). A Global South perspective on explainable AI. *Carnegie Endowment for International Peace*, 30 April. <https://carnegieendowment.org/2024/04/30/global-south-perspective-on-explainable-ai-pub-92333>
- Ehrenfeld, D. (2020). ‘Sharing a world with others’: Rhetoric’s ecological turn and the transformation of the networked public sphere. *Rhetoric Society Quarterly*, 50(5), 305–320. <https://doi.org/10.1080/02773945.2020.1813321>
- Eichler, H. (2023). Cross-Media and Platformised Journalism: ARD’s Innovation Attempts at Becoming a Multi-Platform Contributor to Society. In M. Puppis & C. Ali (Eds), *Public Service Media’s Contribution to Society: RIPE@2021* (pp. 265–288). Nordicom.
- Eisenach, J. A. (2015). *The Economics of Zero Rating*. NERA Economic Consulting, US.
- Elishar-Malka, V., Ariel, Y., & Weimann, G. (2020). Rethinking political communication in the digital sphere. *The Journal of International Communication*, 26(2), 190–210. <https://doi.org/10.1080/13216597.2020.1771397>
- Elkin-Koren, N. (2020). Contesting algorithms: Restoring the public interest in content filtering by artificial intelligence. *Big Data & Society*, 7(2), 1–13. <https://doi.org/10.1177/2053951720932296>
- Elkin-Koren, N., Perel (Filmar), M., & Somech, O. (2024). Unlocking platform data for research. *Proceedings of the TPRC2024, the Research Conference on Communications, Information and Internet Policy*. <https://doi.org/https://doi.org/10.2139/ssrn.4733788>
- Elliott, V. (2022). Meta made millions in ads from networks of fake accounts. *Wired*, 23 June. www.wired.com/story/meta-is-making-millions-from-fake-accounts
- Elliott, V. (2024). AI-fakes detection is failing voters in the Global South. *Wired*, 2 September. www.wired.com/story/generative-ai-detection-gap
- Elmer, G., & Ward-Kimola, S. (2023). Crowdfunding (as) disinformation: ‘Pitching’ 5G and election fraud campaigns on GoFundMe. *Media, Culture & Society*, 45(3), 578–594. <https://doi.org/10.1177/01634437221136009>
- Elsom, C. (2020). *Safety without Censorship: A Better Way to Tackle Online Harms*. Centre for Policy Studies, London.
- Enghel, F., & Noske-Turner, J. (Eds) (2018). *Communication in International Development: Doing Good or Looking Good?* Routledge.
- EP (European Parliament). (2024). *EP Academic Freedom Monitor 2023*. EPRS (European Parliamentary Research Service), STOA (Scientific Foresight Unit), PE 757.798. [www.europarl.europa.eu/thinktank/en/document/EPRS_STU\(2024\)757798](http://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2024)757798)
- Epstein, B. (2020). Why It Is So Difficult to Regulate Disinformation Online. In W. L. Bennett & S. Livingston (Eds), *The Disinformation Age* (pp. 190–210). Cambridge University Press.
- Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences of the United States of America*, 112(33), E4512–E4521. doi: 10.1073/pnas.1419828112.
- ERGA (European Regulators Group for Audiovisual Media Services). (2020). *ERGA Report on Disinformation: Assessment*

of the Implementation of the Code of Practice. <https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf>

- Erickson, J. (2024). Rethinking the filter bubble? Developing a research agenda for the protective filter bubble. *Big Data & Society*, 11(1), 1–4. <https://doi.org/10.1177/20539517241231276>
- Erie, M. S., & Streinz, T. (2021). *The Beijing effect: China's 'digital silk road' as transnational data governance*. New York University Journal of International Law and Politics, 54(1), 1–25. <https://ssrn.com/abstract=3810256>
- Erkmén, Ö. (2023). Data journalism: A systematic literature review. *Journalism Studies*, 25(1), 58–79. <https://doi.org/10.1080/1461670X.2023.2289885>
- Erman, E., & Furendal, M. (2022). The global governance of artificial intelligence: Some normative concerns. *Moral Philosophy and Politics*, 9(2), 267–291. <https://doi.org/10.1515/mopp-2020-0046>
- eSafety Commissioner. (2024a). Report reveals the extent of deep cuts to safety staff and gaps in Twitter/X's measures to tackle online hate. Media Release, 11 January. www.esafety.gov.au/newsroom/media-releases/report-reveals-the-extent-of-deep-cuts-to-safety-staff-and-gaps-in-twitter/xs-measures-to-tackle-online-hate
- eSafety Commissioner. (2024b). *Basic Online Safety Expectations: Summary of response from X Corp. (Twitter) to eSafety's Transparency Notice on Online Hate*. eSafety Commissioner, Australia.
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St Martin's Publishing Group.
- Evanega, S., Lynas, M., Adams, J., & Smolenyak, K. (2022). *Coronavirus Misinformation: Quantifying Sources and Themes in the COVID-19 'Infodemic'*. CALS (Cornell Alliance for Science), US.
- Evans, R., Jackson, D., & Murphy, J. (2023). Google News and machine gatekeepers: Algorithmic personalisation and news diversity in online news search. *Digital Journalism*, 11(9), 1682–1700. <https://doi.org/10.1080/21670811.2022.2055596>
- Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, 22(1), 35–52. <https://doi.org/10.1111/jcc4.12180>
- Faculty. (2021). *Automated Approaches to Measuring Online Experiences: Executive Summary*. www.bleadon.org.uk/media/other/24400/OFCOM-automated-tooling-report-facultyreport.pdf
- Faife, C. (2022). Feds are still using ID.me to scan your face — and human reviewers can't keep up. *The Verge*, 11 February. www.theverge.com/2022/2/11/22928082/id-me-irs-facial-recognition-overworked-employees
- Fallis, D. (2015). What is disinformation? *Library Trends*, 63(3), 401–426. <https://muse.jhu.edu/pub/1/article/579342>
- Farahat, M. (2021). *Egypt Digital Rights Landscape Report*. Institute for Development Studies, Sussex.
- Farkas, J., & Schou, J. (2018). Fake news as a floating signifier: Hegemony, antagonism and the politics of falsehood. *Javnost – The Public*, 25(3), 298–314. <https://doi.org/10.1080/13183222.2018.1463047>
- Farooq, G. (2018). Politics of fake news: How WhatsApp became a potent propaganda tool in India. *Media Watch*, 9(1), 106–117. <https://doi.org/10.15655/mw/2018/v9i1/49279>
- Farrell, H., & Newman, A. (2023). *Underground Empire: How America Weaponized the World Economy*. Allen Lane.
- Fatafta, M. (2024). How Meta censors Palestinian voices. Access Now, 19 February. www.accessnow.org/publication/how-meta-censors-palestinian-voices
- Fawzi, N., Steindl, N., Obermaier, M., Prochazka, F., et al. (2021). Concepts, causes and consequences of trust in news media – A literature review and framework. *Annals of the International Communication Association*, 45(2), 154–174. <https://doi.org/10.1080/23808985.2021.1960181>
- Feathers, T., & Ng, A. (2022). Tech industry groups are watering down attempts at privacy regulation, one state at a time. *The Markup*, 26 May. <https://themarkup.org/privacy/2022/05/26/tech-industry-groups-are-watering-down-attempts-at-privacy-regulation-one-state-at-a-time>
- Fendji, J. L. K. E. (2024). From left behind to left out: Generative AI or the next pain of the unconnected. *Harvard Data Science Review*, 1–3. <https://hdsr.mitpress.mit.edu/pub/10.1177/20539517241231276>
- Fernandez-Monge, F., Barns, S., Bria, F., & Kattel, R. (2024). Reclaiming data for improved city governance: Barcelona's New Data Deal. *Urban Studies*, 61(7), 1291–1307. <https://journals.sagepub.com/doi/abs/10.1177/00420980231204835>
- Ferrara, E. (2024a). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, 6(1), 1–15. <https://doi.org/10.3390/sci6010003>
- Ferrara, E. (2024b). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, 7, 549–569. <https://doi.org/10.1007/s42001-024-00250-1>
- Ferràs-Hernández, X., Nylund, P. A., & Brem, A. (2023). The emergence of dominant designs in Artificial Intelligence. *California Management Review*, 65(3), 73–91. <https://doi.org/10.1177/00081256231164362>
- Fertmann, M., & Kettemann, M. C. (Eds) (2021). *Viral Information: How States and Platforms Deal with Covid-19-Related Disinformation, An Exploratory Study of 20 Countries*. Verlag Hans-Bredow-Institut.
- Feuerriegel, S., DiResta, R., Goldstein, J. A., Kumar, S., et al. (2023). Research can help to tackle AI-generated disinformation. *Nature Human Behaviour*, 7(11), 1818–1821. <https://doi.org/10.1038/s41562-023-01726-2>
- Findlay, M., Ong, L. M., & Zhang, W. (Eds) (2023). *Elgar Companion to Regulating AI and Big Data in Emerging Economies*. Edward Elgar Publishing.

- Finkel, S. E. (2014). The impact of adult civic education programmes in developing democracies. *Public Administration and Development*, 34(3), 169–181. <https://doi.org/10.1002/pad.1678>
- Finkel, S. E., Neundorff, A., & Rascón Ramírez, E. (2024). Can online civic education induce democratic citizenship? Experimental evidence from a new democracy. *American Journal of Political Science*, 68, 613–630. <https://doi.org/10.1111/ajps.12765>
- Fischer-Lescano, A. (2016). Struggles for a global Internet constitution: Protecting global communication structures against surveillance measures. *Global Constitutionalism*, 5(2), 145–172. <https://doi.org/10.1017/S204538171600006X>
- Fitzgerald, C., Williams, K., & Cross, R. J. (2024). *The State of Privacy: How State 'Privacy' Laws Fail to Protect Privacy and How They Can Do Better*. Electronic Privacy Information Center & Public Interest Research Group (PIRG), Education Fund, US.
- Fletcher, R., & Nielsen, R. K. (2024). *What Does the Public in Six Countries Think of Generative AI in News?* Reuters Institute for the Study of Journalism, University of Oxford.
- Flew, T. (2021). *Regulating Platforms*. Polity Press.
- Flew, T. (2023). Mediated trust and artificial intelligence. *Emerging Media*, 1(1), 22–29. <https://doi.org/10.1177/27523543231188793>
- Flew, T., & Martin, F. R. (Eds) (2022). *Digital Platform Regulation: Global Perspectives on Internet Governance*. Springer International.
- Flew, T., & Stepnik, A. (2024). The value of news: Aligning economic and social value from an institutional perspective. *Media and Communication*, 12(2024), 1–16. <https://doi.org/10.17645/mac.7462>
- Flew, T., Iosifidis, P., Meese, J., & Stepnik, A. (2023). Digital platforms and the future of news: Regulating publisher–platform relations in Australia and Canada. *Information, Communication & Society, Online First*, 1–17. <https://doi.org/10.1080/1369118X.2023.2291462>
- Floridi, L. (2013). *The Ethics of Information*. Oxford Academic.
- Floridi, L., & Nobre, A. C. (2024). Anthropomorphising machines and computerising minds: The crosswiring of languages between artificial intelligence and brain & cognitive sciences. *Minds and Machines*, 34(5), 1–9. <https://doi.org/10.2139/ssrn.4738331>
- Fonseka, P. (2024). Catalysing change: Sri Lanka refines digital literacy definition. LIRNEasia, 5 January. <https://lirneasia.net/2024/01/catalysing-change-sri-lanka-refines-digital-literacy-definition>
- Fontes, C., Hohma, E., Corrigan, C. C., & Lürge, C. (2022). AI-powered public surveillance systems: Why we (might) need them and how we want them. *Technology in Society*, 71(2022), 1–12. <https://doi.org/10.1016/j.techsoc.2022.102137>
- Ford, E. (2019). *El Reto de la Democracia Digital: Hacia una Ciudadanía Interconectada*. KAS, JNE, ONPE, D&D Internacional.
- Forum on Information and Democracy. (2023). *Pluralism of News and Information in Curation and Indexing Algorithms: Policy Framework*. February. https://informationdemocracy.org/wp-content/uploads/2023/02/Report_Pluralism-in-algorithms.pdf
- Forum on Information and Democracy. (2024a). *AI as a Public Good: Ensuring Democratic Control of AI in the Information Space*. February. <https://informationdemocracy.org/wp-content/uploads/2024/03/ID-AI-as-a-Public-Good-Feb-2024.pdf>
- Forum on Information and Democracy. (2024b). Data protection authorities stand up to big tech: The importance of democratic institutions to safeguard citizens' privacy rights. Unpacking Current Developments in the Information Space, Insight Nr. 1. <https://informationdemocracy.org/wp-content/uploads/2024/09/FID-Insight-Nr-1-Data-protection-authorities-and-AI.pdf>
- Forum on Information and Democracy. (2024c). The EU's Digital Services Act must provide researchers access to VLOPs' experimental protocols. Press Release on the paper of Alistair Knott, Dino Predreschi, Jonathan Stray & Stuart Russell, 14 June. <https://informationdemocracy.org/2024/06/14/the-forum-members-of-its-working-group-and-researchers-call-for-the-dsa-to-provide-researchers-access-to-conduct-experimental-evaluations-of-vlops>
- Forum on Information and Democracy. (2024d). US Supreme Court Decision on social media regulation, uncertainty remains! Unpacking Current Developments in the Information Space, Insight Nr. 2. <https://informationdemocracy.org/wp-content/uploads/2024/09/FID-Insight-Nr-2-US-Supreme-Court-Decision-on-social-media-regulation.pdf>
- Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings, 1972–1977*. Knopf Doubleday Publishing Group.
- Franco, S. F., Graña, J. M., & Rikap, C. (2024). Dependency in the digital age? The experience of Mercado Libre in Latin America. *Development and Change*, 55(3), 429–464. <https://ideas.repec.org/a/bla/devchg/v55y2024i3p429-464.html>
- François, C. (2019). *Actors, Behaviors, Content: A Disinformation ABC: Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses*. TransAtlantic Working Group, Graphika and Berkman Klein Center for Internet & Society, Harvard University. <https://docs.house.gov/meetings/SY/SY21/20190926/109980/HHRG-116-SY21-Wstate-FrancoisC-20190926-SD001.pdf>
- Franks, M. A. (2021). Beyond the public square: Imagining digital democracy. *Yale Law Journal*, 131(2021–22), 427–453. www.yalelawjournal.org/forum/beyond-the-public-square-imagining-digital-democracy

- Fraser, N. (1992). Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. In C. Calhoun (Ed.), *Habermas and the Public Sphere* (pp. 109–142). MIT Press.
- Fraser, N., & Nash, K. (Eds) (2014). *Transnationalizing the Public Sphere*. Polity Press.
- Frau-Meigs, D. (2022). How disinformation reshaped the relationship between journalism and media and information literacy (MIL): Old and new perspectives revisited. *Digital Journalism*, 10(5), 912–922. <https://doi.org/10.1080/21670811.2022.2081863>
- Frau-Meigs, D. (2024a). Algorithm literacy as a subset of media and information literacy: Competences and design considerations. *Digital*, 4(2), 512–528. <https://doi.org/10.3390/digital4020026>
- Frau-Meigs, D. (2024b). On AI Pedagogy. Presentation to VOICES – European Festival of Journalism and Media Literacy, March.
- Frau-Meigs, D. (2024c). *User Empowerment through Media and Information Literacy Responses to the Evolution of Generative Artificial Intelligence (GAI)*. Report prepared for UNESCO, Sorbonne Nouvelle University.
- Frau-Meigs, D., & Corbu, N. (Eds) (2024). *Disinformation Debunked: Building Resilience through Media and Information Literacy*. Routledge.
- Frau-Meigs, D., Velez, I., & Michel, J. F. (Eds) (2017). *Public Policies in Media and Information Literacy in Europe: Cross-Country Comparisons*. Routledge.
- Fredheim, R., Bay, S., Dek, A., Stolze, M., & Haiduchyk, T. (2023). *Social Media Manipulation 2022/2023: Assessing the Ability of Social Media Companies to Combat Platform Manipulation*. NATO Strategic Communications Centre of Excellence. <https://stratcomcoe.org/publications/social-media-manipulation-20222023-assessing-the-ability-of-social-media-companies-to-combat-platform-manipulation/272>
- Freedman, D. (2014). The puzzle of media power: Notes toward a materialist approach. *International Journal of Communication*, 8(2014), 319–334. <https://ijoc.org/index.php/ijoc/article/view/2081>
- Freedom House. (2024). *The Struggle for Trust Online*. Freedom on the Net 2024. <https://freedomhouse.org/report/freedom-net/2024/struggle-trust-online>
- Freedom Online Coalition. (2023). Aims and priorities. <https://freedomonlinecoalition.com/aims-and-priorities>
- Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political Communication*, 37(2), 145–156. <https://doi.org/10.1080/10584609.2020.1723755>
- Freelon, D., Bossetta, M., Wells, C., Lukito, J., Xia, Y., & Adams, K. (2022). Black trolls matter: Racial and ideological asymmetries in social media disinformation. *Social Science Computer Review*, 40(3), 560–578. <https://doi.org/10.1177/0894439320914853>
- Freire, P. (1974). *Education for Critical Consciousness*. Continuum.
- Frieden, R. (2014). *Ex Ante Versus Ex Post Approaches to Network Neutrality: A Cost Benefit Analysis*. <http://dx.doi.org/10.2139/ssrn.2493945>
- Friesem, Y., Raman, U., Kanižaj, I., & Choi, G. Y. (Eds) (2022). *The Routledge Handbook of Media Education Futures Post-Pandemic*. Routledge.
- Frislich, L., & Humprecht, E. (2021). *Trust, Democratic Resilience, and the Infodemic*. Policy Paper. Israel Public Policy Institute, Heinrich Böll Stiftung, & Embassy of the Federal Republic of Germany.
- FTC (Federal Trade Commission). (2024). FTC launches inquiry into generative AI investments and partnerships. Press Release, 25 January. www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships
- Fuchs, C. (2023). *Digital Democracy and the Digital Public Sphere: Media Communication and Society, Volume Six*. Taylor & Francis.
- Fung, A., & Wright, E. O. (Eds) (2003). *Deepening Democracy: Institutional Innovations in Empowered Participatory Governance*. Verso.
- Funk, A., Shahbaz, A., & Vesteinsson, K. (2023). Freedom on the Net 2023: *The Repressive Power of Artificial Intelligence*. Freedom House. <https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence>
- Furstenberg, S., Prelic, T., & Heathershaw, J. (2020). *The Internationalization of Universities and the Repression of Academic Freedom*. Freedom House US. <https://freedomhouse.org/report/special-report/2020/internationalization-universities-and-repression-academic-freedom>
- G20. (2019). *G20 Ministerial Statement on Trade and Digital Economy, Annex G20 AI Principles*. <https://wp.oecd.ai/app/uploads/2021/06/G20-AI-Principles.pdf>
- G7. (2023). Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems. G7 2023 Hiroshima Summit. <https://www.mofa.go.jp/files/100573471.pdf>
- Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering Online Hate Speech*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000233231>
- Galantino, S. (2023). How will the EU Digital Services Act affect the regulation of disinformation? *SCRIPTed: A Journal of Law, Technology and Society*, 20(1), 89–129. <https://script-ed.org/wp-content/uploads/2023/02/Sharon-Galantino-February-2023.pdf?d=05112024>

- Galič, M. (2022). Smart cities as 'Big Brother only to the masses': The limits of personal privacy and personal surveillance. *Surveillance & Society*, 20(3), 306–311. <https://doi.org/10.2139/ssrn.4228444>
- Gallagher, A., O'Connor, C., & Visser, F. (2023). *Uisce Faoi Thalamh: Summary Report – An Investigation into the Online Mis- and Disinformation Ecosystem in Ireland*. ISD (Institute for Strategic Dialogue).
- Gallagher, M. (2023). Gendered Disinformation and Platform Accountability. In M. Gallagher & A. Vega Montiel (Eds), *The Handbook of Gender, Communication, and Women's Human Rights* (pp. 53–69). John Wiley & Sons.
- Galli, F., Loreggia, A., & Sartor, G. (2023). The Regulation of Content Moderation. In D. Moura Vicente, S. de Vasconcelos Casimiro, & C. Chen (Eds), *The Legal Challenges of the Fourth Industrial Revolution* (pp. 63–87). Springer International Publishing.
- Garcia Ramirez, D. (2021). Journalism in the attention economy – The relation between digital platforms and news organisations. *Brazilian Journalism Research*, 17(1), 1–24. <https://doi.org/10.25200/BJR.v17n1.2021.1332>
- Garcia Santamaria, S., & Salojärvi, V. (2020). Media in Authoritarian Contexts: A Logics Approach to Journalistic Professional Resistance in Cuba and Venezuela. In X. Orchard, S. Garcia Santamaria, J. Brambila, & J. Lugo-Ocando (Eds), *Media & Governance in Latin America: Towards Plurality of Voices* (pp. 97–116). Peter Lang.
- Garnett, H. A., James, T. S., & Caal-Lam, S. (2024). *The Electoral Integrity Global Report 2024*. The Electoral Integrity Project at Royal Military College of Canada, Queen's University, Canada and University of East Anglia UK.
- Gates, B. (2023). The age of AI has begun. GatesNotes Blog, 21 March. www.gatesnotes.com/The-Age-of-AI-Has-Begun
- Gautier, A., & Lamesch, J. (2021). *Mergers in the digital economy. Information Economics and Policy*, 54(2021), 1–15. <https://doi.org/10.1016/j.infoecopol.2020.100890>
- Geiger, C. (2024). Elaborating a human rights-friendly copyright framework for generative AI. *IIC – International Review of Intellectual Property and Competition Law*, 2024(3 June), 1–37. <https://doi.org/10.1007/s40319-024-01481-5>
- Geiß, S., Magin, M., Jürgens, P., & Stark, B. (2021). Loopholes in the echo chambers: How the echo chamber metaphor oversimplifies the effects of information gateways on opinion expression. *Digital Journalism*, 9(5), 660–686. <https://doi.org/10.1080/21670811.2021.1873811>
- Georgiou, M. (2023). *Being Human in Digital Cities*. Polity Press.
- Gerbaudo, P. (2022). Theorizing reactive democracy: The social media public sphere, online crowds, and the plebiscitary logic of online reactions. *Democratic Theory*, 9(2), 120–138. <https://doi.org/10.3167/dt.2022.090207>
- Gerbner, G., Gross, L., Signorielli, N., & Morgan, M. (1980). Aging with television: Images on television drama and conceptions of social reality. *Journal of Communication*, 30(1), 37–47. <https://eric.ed.gov/?id=EJ221287>
- Germain, T. (2024). X bans and then unbans journalists and podcasters in Twitter's latest free speech massacre. *Gizmodo*, 9 January. <https://gizmodo.com/elon-musk-x-twitter-journalists-banning-spree-1851151593>
- German Law Archive. (2017). *Network Enforcement Act (Netzdurchsetzungsgesetz, NetzDG)*. Government of Germany.
- Gerpott, T. J. (2018). Zero-rating arrangements of mobile Internet access service providers – An analysis of main factors shaping the need for regulatory interventions. *Telecommunications Policy*, 42(6), 489–500. <https://doi.org/10.1016/j.telpol.2018.03.003>
- Ghosal, S. S., Chakraborty, S., Geiping, J., Huang, F., Manocha, D., & Bedi, A. S. (2023). Towards possibilities & impossibilities of AI-generated text detection: A survey. *ArXiv*. <https://doi.org/10.48550/arXiv.2310.15264>
- Ghose, R. (2001). Use of information technology for community empowerment: Transforming geographic information systems into community information systems. *Transactions in GIS*, 5(2), 141–163. <https://doi.org/10.1111/1467-9671.00073>
- Ghosh, D., & Scott, B. (2018). *Digital Deceit: The Technologies Behind Precision Propaganda on the Internet*. New America, Public Interest Technology, & Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, US.
- Gibson, A. D., Docherty, N., & Gillespie, T. (2023). Health and toxicity in content moderation: The discursive work of justification. *Information, Communication & Society*, 27(7), 1441–1457. <https://doi.org/10.1080/1369118X.2023.2291456>
- Gil de Zúñiga, H., & Goyanes, M. (2023). Fueling civil disobedience in democracy: WhatsApp news use, political knowledge, and illegal political protest. *New Media & Society*, 25(12), 3500–3520. <https://doi.org/10.1177/14614448211047850>
- Gil de Zúñiga, H., Borah, P., & Goyanes, M. (2021). How do people learn about politics when inadvertently exposed to news? Incidental news paradoxical direct and indirect effects on political knowledge. *Computers in Human Behavior*, 121, 1–9. www.sciencedirect.com/science/article/pii/S0747563221001266?casa_token=eOjAerFI9RkAAAAA:KkK-AdlMxQB9UvbkQcaAsRNknfBOZP_COAOYXkqV8j6satCJLfeKJoThF7xGxmvMxVH2Tf
- Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. (2023). A scholarly definition of artificial intelligence (AI): Advancing AI as a conceptual framework in communication research. *Political Communication*, 41(2), 317–334. <https://doi.org/10.1080/10584609.2023.2290497>
- Gillespie, T. (2010). The politics of 'platforms'. *New Media & Society*, 12(3), 347–364. <https://doi.org/10.1177/1461444809342738>
- Gillespie, T. (2020). Content moderation, AI, and the question of scale. *Big Data & Society*, 7(2), 1–5. <https://doi.org/10.1177/2053951720943234>
- Gillespie, T., Aufderheide, P., Carmi, E., Gerrard, Y., et al. (2023). Expanding the debate about content moderation: Scholarly research agendas for the coming policy debates. *Internet Policy Review*, 9(4). <https://ssrn.com/abstract=4459448>

- Gillwald, A. (2024). Harnessing global cooperation: Africa's role in shaping the digital future after the Global Digital Compact. *Research ICT Africa*. <https://mailchi.mp/researchictafrica/africas-role-in-shaping-the-digital-future-after-the-global-digital-compact?e=43ad1c2882>
- Gillwald, A., & Wavre, V. (2024). Rerouting Geopolitics: Narratives and the Political Power of Communication. In C. Padovani, V. Wavre, A. Hintz, G. Goggin, & P. Iosifidis (Eds), *Global Communication Governance at the Crossroads* (pp. 21–35). Palgrave Macmillan.
- Gillwald, A., & Weleilakeba, S. (2024). *From the Digital Inequality Paradox to Marginalisation through Digitalisation: GiSWatch 2024 Special Edition reports*. Association for Progressive Communications.
- Gillwald, A., Murkic, D., Leslie, D., Katell, M., et al. (2022). *Data Justice in Practice: A Guide for Impacted Communities*. Global Partnership on Artificial Intelligence.
- Gitelman, L. (2006). *Always Already New: Media, History, and the Data of Culture*. MIT Press.
- Gitelman, L. (Ed.) (2013). *'Raw Data' is an Oxymoron*. MIT Press.
- Giusti, S., & Piras, E. (Eds) (2021). *Democracy and Fake News: Information Manipulation and Post-Truth Politics*. Routledge.
- Glasberg, D. S., & Shannon, D. (2010). *Political Sociology: Oppression, Resistance, and the State*. SAGE Publications.
- Glimmerveen, L., Ybema, S., & Nies, H. (2022). Who participates in public participation? The exclusionary effects of inclusionary efforts. *Administration & Society*, 54(4), 543–574. <https://doi.org/10.1177/00953997211034137>
- Global Kids Online. (2019). *Global Kids Online: Comparative Report*. UNICEF Office of Research – Innocenti. www.unicef.org/innocenti/media/7011/file/GKO-Comparative-Report-2019.pdf
- Global Witness. (2023). How big tech platforms are neglecting their non-English language users. 30 November. www.globalwitness.org/en/campaigns/digital-threats/how-big-tech-platforms-are-neglecting-their-non-english-language-users
- Goffman, E. (1974). *Frame Analysis: An Essay on the Organization of Experience*. Harvard University Press.
- Goldman Sachs. (2024). AI is poised to drive 160% increase in data center power demand. Goldman Sachs. 14 May. <https://www.goldmansachs.com/insights/articles/AI-poised-to-drive-160-increase-in-power-demand>
- Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., & Tomz, M. (2024). How persuasive is AI-generated propaganda? *Proceedings of the National Academy of Sciences Nexus*, 3(2), 1–7. <https://doi.org/10.1093/pnasnexus/pgae034>
- Gollwitzer, A., Martel, C., Brady, W. J., Pärnamets, P., et al. (2020). Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. *Nature Human Behaviour*, 4(11), 1186–1197. <https://doi.org/10.1038/s41562-020-00977-7>
- Gonçalves, J., Weber, I., Masullo, G. M., Torres da Silva, M., & Hofhuis, J. (2023). Common sense or censorship: How algorithmic moderators and message type influence perceptions of online content deletion. *New Media & Society*, 25(10), 2595–2617. <https://doi.org/10.1177/14614448211032310>
- González-Bailón, S., Lazer, D., Barberá, P., Zhang, M., et al. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. *Science*, 381(6656), 392–398. <https://doi.org/10.1126/science.ade7138>
- Google. (2022). Our commitment to climate-conscious data center cooling. 21 November. <https://blog.google/outreach-initiatives/sustainability/our-commitment-to-climate-conscious-data-center-cooling>
- Gordon-Tapiero, A., Ohm, P., & Ramaswami, A. (2023). Fact and Friction: A Case Study in the Fight Against False News. *UC Davis Law Review*, 57, November. <https://lawreview.law.ucdavis.edu/archives/57/1/fact-and-friction-case-study-fight-against-false-news>
- Gorwa, R., & Guilbeault, D. (2020). Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet*, 12(2), 225–248. <https://doi.org/10.1002/poi3.184>
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 1–15. <https://doi.org/10.1177/2053951719897945>
- Gotfredsen, S. G., & Dowling, K. (2024). Meta is getting rid of CrowdTangle – And its replacement isn't as transparent or accessible. *Columbia Journalism Review*, 9 July. www.cjr.org/tow_center/meta-is-getting-rid-of-crowdtangle.php
- Government of Brazil. (2018). *Lei Geral de Proteção de Dados Pessoais (LGPD)*.
- Government of Brazil. (2023). *Projeto de Lei 2338/2023, Dispõe sobre o uso da Inteligência Artificial*. Internal Temporary Commission on Artificial Intelligence in Brazil.
- Government of France. (2018). *Regarding the Fight Against Information Manipulation*. Organic Law No. 2018–1201 of 22 December 2018.
- Government of India. (2023). The Digital Personal Data Protection Act, 2023. Ministry of Electronics & Information Technology. <https://www.meity.gov.in/content/digital-personal-data-protection-act-2023>
- Government of India. (2023). *The Digital Personal Data Protection Act, 2023*.
- Graham, M., & Dittus, M. (2022). *Geographies of Digital Exclusion: Data and Inequality*. Pluto Press.
- Graham, M., & Ferrari, F. (Eds) (2022). *Digital Work in the Planetary Market*. MIT Press.
- Graham, T. (2024). Is big tech harming society? To find out, we need research – but it's being manipulated by big tech itself. *The Conversation*, 3 October. <https://theconversation.com/is-big-tech-harming-society-to-find-out-we-need-research-but-its-being-manipulated-by-big-tech-itself-240110>

- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Graves, L., & Cherubini, F. (2016). *The Rise of Fact-Checking Sites in Europe*. Reuters Institute for the Study of Journalism, University of Oxford.
- Green, B. (2022). Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. *Philosophy & Technology*, 35(4), 1–32. <https://doi.org/10.1007/s13347-022-00584-6>
- Green, B., & Kak, A. (2021). The false comfort of human oversight as an antidote to AI harm. *Slate*, 15 June. <https://ssrn.com/abstract=4046163>
- Greenberg, W. (2022). *How ad tech became cop spy tech*. Electronic Frontier Foundation, 31 August. www.eff.org/deeplinks/2022/08/how-ad-tech-became-cop-spy-tech
- Greene, L. (2018). *Silicon States: The Power and Politics of Big Tech and What It Means for Our Future*. Counterpoint.
- Greenstein, S., Peitz, M., & Valletti, T. (2016). Net neutrality: A fast lane to understanding the trade-offs. *Journal of Economic Perspectives*, 30(2), 127–150. <https://doi.org/10.1257/jep.30.2.127>
- Gritsenko, D., Markham, A., Pötzsch, H., & Wijermars, M. (2022). Algorithms, contexts, governance: An introduction to the special issue. *New Media & Society*, 24(4), 835–844. <https://doi.org/10.1177/14614448221079037>
- Grizzle, A., Wilson, C., Tuazon, R., Cheung, C. K., et al. (2021). *Media and Information Literate Citizens: Think Critically, Click Wisely!* UNESCO.
- Grover, R., & Baik, J. (2024). Platforms as templates: Emerging datafication dynamics in digital news outlets' datawalls. *The Information Society*, 40(4), 260–276. <https://doi.org/10.1080/01972243.2024.2350015>
- Guallar, J., Codina, L., Freixa, P., & Pérez-Montoro, M. (2022). Disinformation, hoaxes, curation and verification: Review of studies in Ibero-America 2017–2020. *Online Media and Global Communication*, 1(3), 648–668. <https://doi.org/10.1515/omgc-2022-0055>
- Guess, A. M., Malhotra, N., Pan, J., Barbera, P., et al. (2023a). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656), 398–404. www.science.org/doi/10.1126/science.abp9364
- Guess, A. M., Malhotra, N., Pan, J., Barbera, P., et al. (2023b). Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science*, 381(6656), 404–408. www.science.org/doi/10.1126/science.add8424
- Guess, A. M., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), 1–8. <https://doi.org/10.1126/sciadv.aau4586>
- Gulati, R. (2023). Meta's Oversight Board and transnational hybrid adjudication – What consequences for international law? *German Law Journal*, 24(3), 473–493. <https://doi.org/10.1017/glj.2023.34>
- Gunkel, D. J. (Ed.). (2024). *Handbook on the Ethics of Artificial Intelligence*. Edward Elgar Publishing.
- Gunkel, D. J., Ali, S. M., Paragi, B., Daly, A. C., et al. (2024). The (Un)bearable Whiteness of AI Ethics. In D. J. Gunkel (Ed.), *Handbook on the Ethics of Artificial Intelligence* (pp. 218–231). Edward Elgar Publishing.
- Gunton, K. (2022). The Impact of the Internet and Social Media Platforms on Radicalisation to Terrorism and Violent Extremism. In R. Montasari, F. Carroll, I. Mitchell, S. Hara, & R. Bolton-King (Eds), *Privacy, Security and Forensics in The Internet of Things (IoT)* (pp. 167–177). Springer International Publishing.
- Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics*, 10, 178–206. https://doi.org/10.1162/tacl_a_00454
- Guriev, S., & Treisman, D. (2019). Informational autocrats. *The Journal of Economic Perspectives*, 33(4), 100–127. www.jstor.org/stable/26796838
- Guriev, S., Henry, E., Marquis, T., & Zhuravskaya, E. (2023). Curtailing false news, amplifying truth. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4616553
- Gurumurthy, A., & Bharthur, D. (2023). *Reframing AI Governance through a Political Economy Lens*. IT for Change. <https://itforchange.net/index.php/reframing-ai-governance-through-a-political-economy-lens>
- Gurumurthy, A., & Chami, N. (2019). *The Wicked Problem of AI Governance*. IT for Change, India and Friedrich Ebert Stiftung.
- Gurumurthy, A., & Chami, N. (2022). *Governing the Resource of Data: To What End and for Whom? Conceptual Building Blocks of a Semi-Commons Approach*. IDFC Institute Working Paper No. 23. Data Governance Network, International Development Finance Club.
- Gurumurthy, A., & Chami, N. (2024). Justice in the Post-Public Sphere: The New Challenge for Global Governance. In C. Padovani, V. Wavre, A. Hintz, G. Goggin, & P. Iosifidis (Eds), *Global Communication Governance at the Crossroads* (pp. 111–127). Palgrave Macmillan.
- Gutiérrez-Coba, L. M., Coba-Gutiérrez, P., & Gómez-Díaz, J. A. (2020). Fake news about Covid-19: A comparative analysis of six Iberoamerican countries. *Revista Latina de Comunicación Social*, 78(2020), 237–264. <https://doi.org/10.4185/RLCS-2020-1476>
- Ha, S., Park, Y., Kim, J., & Kim, S. (2023). Research trends of digital platforms: A survey of the literature from 2018 to 2021. *Telecommunications Policy*, 47(8), 1–16. <https://doi.org/10.1016/j.telpol.2023.102543>
- Haas, J., & Kettemann, M. (2024). *Platform and Content Governance in Times of Crisis*. OSCE (Organization for Security and Co-operation in Europe).

- Habermas, J. (2015). *The Structural Transformation of the Public Sphere: An Inquiry Into a Category of Bourgeois Society*. John Wiley & Sons [first published in 1962 in German; first English translation 1989].
- Habermas, J. (2022). Reflections and hypotheses on a further structural transformation of the political public sphere. *Theory, Culture & Society*, 39(4), 145–171. <https://doi.org/10.1177/02632764221112341>
- Hadi Mogavi, R., Deng, C., Juho Kim, J., Zhou, P., et al. (2024). ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior: Artificial Humans*, 2(1), 1–20. <https://doi.org/10.1016/j.chbah.2023.100027>
- Haenig, M. A., & Ji, X. (2024). A tale of two Southeast Asian states: Media governance and authoritarian regimes in Singapore and Vietnam. *Asian Review of Political Economy*, 3(4), 1–23. <https://doi.org/10.1007/s44216-024-00024-6>
- Haider, J., & Sundin, O. (2022). *Paradoxes of Media and Information Literacy: The Crisis of Information*. Taylor & Francis.
- Haidt, J., & Bail, C. (2024). *Social Media and Political Dysfunction: A Collaborative Review*. New York University.
- Hall, P., & Ellis, D. (2023). A systematic review of socio-technical gender bias in AI algorithms. *Online Information Review*, 47(7), 1264–1279. <https://doi.org/10.1108/OIR-08-2021-0452>
- Hallin, D. C., & Mancini, P. (2004). *Comparing Media Systems: Three Models of Media and Politics*. Cambridge University Press.
- Hallin, D. C., & Mancini, P. (Eds) (2012). *Comparing Media Systems Beyond the Western World*. Cambridge University Press.
- Halpern, H., Brown, G., Kagan, O., Li, L., & Avramenko, A. (2024). Google accused of privacy breaches over alleged Chrome tracking. GRC World Forums, 14 June. www.grcworldforums.com/risk/google-accused-of-privacy-breaches-over-alleged-chrome-tracking/9675.article
- Hameleers, M., & Marquart, F. (2023). It's nothing but a deepfake! The effects of misinformation and deepfake labels delegitimizing an authentic political speech. *International Journal of Communication*, 17(2023), 6291–6307. <https://ijoc.org/index.php/ijoc/article/view/20777>
- Hamelink, C. J. (2023). *Communication and Human Rights: Towards Communicative Justice*. Polity.
- Han, T. A., Lenaerts, T., Santos, F. C., & Pereira, L. M. (2022). Voluntary safety commitments provide an escape from over-regulation in AI development. *Technology in Society*, 68(2022), 1–14. <https://doi.org/10.1016/j.techsoc.2021.101843>
- Hancock, E. (2024). The EU's uphill battle against Big Tech power. *POLITICO*, 6 March. www.politico.eu/article/the-eus-uphill-battle-against-big-tech-power
- Hanitzsch, T., & Vos, T. P. (2018). Journalism beyond democracy: A new look into journalistic roles in political and everyday life. *Journalism*, 19(2), 146–164. <https://doi.org/10.1177/1464884916673386>
- Hanitzsch, T., Van Dalen, A., & Steindl, N. (2018). Caught in the nexus: A comparative and longitudinal analysis of public trust in the press. *The International Journal of Press/Politics*, 23(1), 3–23. <https://doi.org/10.1177/1940161217740695>
- Hao, K. (2020). We read the paper that forced Timnit Gebru out of Google. Here's what it says. MIT Technology Review, 4 December. www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru
- Hao, K. (2022). A new vision of artificial intelligence for the people. *MIT Technology Review*, 22 April. www.technologyreview.com/2022/04/22/1050394/artificial-intelligence-for-the-people
- Hao, X., Demir, E., & Eyers, D. (2024). Exploring collaborative decision-making: A quasi-experimental study of human and Generative AI interaction. *Technology in Society*, 78(2024), 1–22. <https://doi.org/10.1016/j.techsoc.2024.102662>
- Harbath, K. (2023). Different approaches to counting elections. Anchor Change, 19 December. <https://anchorchange.substack.com/p/different-approaches-to-counting>
- Hardy, C., & Williams, S. (2008). E-government policy and practice: A theoretical and empirical exploration of public e-procurement. *Government Information Quarterly*, 25(2), 155–180. <https://doi.org/10.1016/j.giq.2007.02.003>
- Hargittai, E. (Ed.) (2021). *Handbook of Digital Inequality*. Edward Elgar Publishing.
- Harlow, S. (2017). *Liberation Technology in El Salvador: Re-appropriating Social Media among Alternative Media Projects*. Springer.
- Harlow, S. (2022). A new people's press? Understanding digital-native news sites in Latin America as alternative media. *Digital Journalism*, 10(8), 1322–1341. <https://doi.org/10.1080/21670811.2021.1907204>
- Harrington, E., & Vermeulen, M. (2024). *External Researcher Access to Closed Foundation Models: State of the Field and Options for Improvement*. AWO supported by the Mozilla Foundation. <https://blog.mozilla.org/wp-content/blogs.dir/278/files/2024/10/External-researcher-access-to-closed-foundation-models.pdf>
- Harris, B. (2023). Google draws backlash from Brazil with lobbying against 'fake news' bill. *The Financial Times*, 2 May. www.ft.com/content/8a4bb131-e792-491d-a348-9dee4de41ce2
- Harris, J. (2023). 'There was all sorts of toxic behaviour': Timnit Gebru on her sacking by Google, AI's dangers and big tech's biases. *The Guardian*, 22 May. www.theguardian.com/lifeandstyle/2023/may/22/there-was-all-sorts-of-toxic-behaviour-timnit-gebru-on-her-sacking-by-google-ais-dangers-and-big-techs-biases
- Harsin, J. (2024). Three critiques of disinformation (for-hire) scholarship: Definitional vortexes, disciplinary unneighborliness, and cryptonormativity. *Social Media + Society*, 10(1), 1–12. <https://doi.org/10.1177/20563051231224732>
- Hartley, J. M., Petre, C., Bengtsson, M., & Kammer, A. (2023). Autonomies and dependencies: Shifting configurations of power in the platformization of news. *Digital Journalism*, 11(8), 1375–1390. <https://doi.org/10.1080/21670811.2023.2257759>

- Harvie, D., Lightfoot, G., Lilley, S., & Weir, K. (2013). *Publisher, be damned! From price gouging to the open road*. Prometheus, 31(3), 229–239. <https://doi.org/10.1080/08109028.2014.891710>
- Harwell, D. (2019). FBI, ICE find state driver's license photos are a gold mine for facial-recognition searches. *The Washington Post*, 7 July. www.washingtonpost.com/technology/2019/07/07/fbi-ice-find-state-drivers-license-photos-are-gold-mine-facial-recognition-searches
- Harwell, D. (2021). ICE investigators used a private utility database covering millions to pursue immigration violations. *The Washington Post*, 26 February. www.washingtonpost.com/technology/2021/02/26/ice-private-utility-data
- Hase, V., Boczek, K., & Scharrow, M. (2023). Adapting to affordances and audiences? A cross-platform, multi-modal analysis of the platformization of news on Facebook, Instagram, TikTok, and Twitter. *Digital Journalism*, 11(8), 1499–1520. <https://doi.org/10.1080/21670811.2022.2128389>
- Hasimi, L., & Poniszewska-Marañda, A. (2024). Detection of disinformation and content filtering using machine learning: Implications to human rights and freedom of speech. ROMCIR 2024: The 4th Workshop on Reducing Online Misinformation through Credible Information Retrieval (held as part of ECIR 2024, the 46th European Conference on Information Retrieval). <https://ceur-ws.org/Vol-3677/paper6.pdf>
- Hatfield, A., Meghiani, K., Williams-Bellamy, J., Saiyid, D., & Zhou, J. (2023). *Mapping and Recommendations for UNDP Efforts to Fight Against Mis- and Disinformation*. Columbia SIPA and UNDP Accelerator Labs.
- He, A. (2023). *State-Centric Data Governance in China*. CIGI Papers No. 282. CIGI (Centre for International Governance Innovation), Canada.
- Heeks, R. (2022). Digital inequality beyond the digital divide: Conceptualizing adverse digital incorporation in the global South. *Information Technology for Development*, 28(4), 688–704. <https://doi.org/10.1080/02681102.2022.2068492>
- Heitz, L., Rozgonyi, K., & Kostic, B. (2021). AI in Content Curation and Media Pluralism. In D. Wagner & J. Haas (Eds), *Spotlight on Artificial Intelligence and Freedom of Expression – A Policy Manual* (pp. 56–70). OSCE (Organization for Security and Co-operation in Europe).
- Helberger, N., Karppinen, K., & D'Acunzio, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, 21(2), 191–207. <https://doi.org/10.1080/1369118X.2016.1271900>
- Helberger, N., Van Drunen, M., Eskens, S., & Bastian, M. (2020). A freedom of expression perspective on AI in the media – with a special focus on editorial decision making on social media platforms and in the news media. *European Journal of Law and Technology*, 11(3), 1–28. <https://ejlt.org/index.php/ejlt/article/view/752>
- Helberger, N., Van Drunen, M., Moeller, J., Vrijenhoek, S., & Eskens, S. (2022). Towards a normative perspective on journalistic AI: Embracing the messy reality of normative ideals. *Digital Journalism*, 10(10), 1605–1626. <https://doi.org/10.1080/21670811.2022.2152195>
- Helios-i.mashable.com. (2024). X image. https://helios-i.mashable.com/imagery/articles/OOITPG6lvAMMP2M4bIHAvwX/hero-image.fill.size_1248x702.v1725051736.jpg
- Helming, C. (2023). Microsoft's Bing Chat: A source of misinformation on elections. *Algorithm Watch*, 15 December. <https://algorithmwatch.org/en/microsofts-bing-source-misinformation-elections>
- Helsper, E. J. (2021). *The Digital Disconnect: The Social Causes and Consequences of Digital Inequalities*. SAGE Publications.
- Henderson, J., & Corry, M. (2020). Data literacy training and use for educational professionals. *Journal of Research in Innovative Teaching & Learning*, 14(2), 232–244. <https://doi.org/10.1108/JRIT-11-2019-0074>
- Heng, S., Tsilonis, K., Scharff, C., & Wautelet, Y. (2022). Understanding AI ecosystems in the Global South: The cases of Senegal and Cambodia. *International Journal of Information Management*, 64(2022), 1–18. <https://doi.org/10.1016/j.ijinfomgt.2021.102454>
- Henningsen, A., & Krčál, A. (2024). *Public Financing of News Media in the EU: Final Report*. Henningsen Consulting and Technopolis.
- Hepp, A., Jarke, J., & Kramp, L. (2022). New Perspectives in Critical Data Studies: The Ambivalences of Data Power – An Introduction. In A. Hepp, J. Jarke, & L. Kramp (Eds), *New Perspectives in Critical Data Studies: The Ambivalences of Data Power* (pp. 1–23). Springer International Publishing.
- Herman, E. S., & Chomsky, N. (1989). *Manufacturing Consent: The Political Economy of the Mass Media*. Penguin.
- Hermida, A. (2023). Journalism prepares for a post-search, post-social future. Nieman Lab. www.niemanlab.org/2023/12/journalism-prepares-for-a-post-search-post-social-future
- Higgins, J., Thomas, J., Chandler, J., Cumpston, M., et al. (Eds) (2024). *Cochrane Handbook for Systematic Reviews of Interventions, Version 6.5*. Cochrane Training.
- Hildebrandt, C., & Wiewiorra, L. (2024). The past, present, and future of (net) neutrality: A state of knowledge review and research agenda. *Journal of Information Technology*, 39(1), 167–193. <https://doi.org/10.1177/02683962231170891>
- Himma-Kadakas, M., & Ojams, I. (2022). Debunking false information: Investigating journalists' fact-checking skills. *Digital Journalism*, 10(5), 866–887. <https://doi.org/10.1080/21670811.2022.2043173>
- Hinds, J., Williams, E. J., & Joinson, A. N. (2020). 'It wouldn't happen to me': Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*, 143(2020), 1–55. <https://doi.org/10.1016/j.ijhcs.2020.102498>

- Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2019). *Digital Citizenship in a Datafied Society*. Polity Press.
- Hlomani, H. (2023). Future-proofing African democracy: Assessing electoral laws amidst AI advancements. Research ICT Africa, Blog, 12 September. <https://researchictafrica.net/2023/09/12/future-proofing-african-democracy-assessing-electoral-laws-amidst-ai-advancements>
- Hofmann, J. (2024). Desinformation als Symptom. Ein Überblick. In *Zentrum für Analyse/Forschung am Bundesamt für Verfassungsschutz (Hg.): Wissenschaftskonferenz 2023. Meinungsbildung 2.0 – Strategien im Ringen um Deutungshoheit im digitalen Zeitalter* (pp. 20–31). Bundesamt für Verfassungsschutz.
- Holt, J. (2023). Information disorder in the Israel–Hamas war highlights shifts in the fight against online misinformation. ISD Digital Dispatches, 9 November. www.isdglobal.org/digital_dispatches/information-disorder-in-the-israel-hamas-war-highlights-shifts-in-the-fight-against-online-misinformation
- Hooper, L., Livingstone, S., & Pothong, K. (2022). *Problems with Data Governance in UK Schools: The Cases of Google Classroom and Class Dojo*. Digital Futures Commission, 5Rights Foundation. <https://digitalfuturescommission.org.uk/wp-content/uploads/2022/08/Problems-with-data-governance-in-UK-schools.pdf>
- Hopkins, J. (2020). The Concept of Affordances in Digital Media. In H. Friese, M. Nolden, G. Rebane, & M. Schreiter (Eds), *Handbuch Soziale Praktiken und Digitale Alltagswelten* (pp. 47–54). Springer VS.
- Horowitz, M., Milosavljević, M., & Van den Bulck, H. (2022). The Use of Artificial Intelligence by Public Service Media: Between Advantages and Threats. In C. El Morr (Ed.), *AI and Society: Tensions and Opportunities* (pp. 14). Chapman and Hall/CRC.
- Horowitz, M., Nieminen, H., Lehtisaari, K., & D’Arma, A. (Eds) (2024). *Epistemic Rights in the Era of Digital Disruption*. Springer International Publishing.
- Horst, H. A., Sargent, A., & Gaspard, I. (2024). Beyond extraction: Data strategies from the Global South. *New Media & Society*, 26(3), 1366–1383. <https://doi.org/10.1177/14614448231201651>
- Horten, M. (2023). Algorithms patrolling content: Where’s the harm? *International Review of Law, Computers & Technology*, 38(1), 43–65. <https://doi.org/10.1080/13600869.2023.2221823>
- Horwitz, J. (2021). The Facebook files. *Wall Street Journal*. www.wsj.com/articles/the-facebook-files-11631713039
- Hoskins, G. T. (2024). Far right > digital rights: The precarity of free expression, internet access, net neutrality and data privacy in Bolsonaro’s Brazil. *Javnost – The Public*, 31(2), 309–326. <https://doi.org/10.1080/13183222.2024.2346704>
- Houser, K., & Voss, W. G. (2018). GDPR: The end of Google and Facebook or a new paradigm in data privacy? *Richmond Journal of Law & Technology*, 25(1), 1–109. https://jolt.richmond.edu/files/2018/11/Houser_Voss-FE.pdf
- Howard, P. N., & Hussain, M. M. (2013). *Democracy’s Fourth Wave? Digital Media and the Arab Spring*. Oxford University Press.
- Howard, P. N., Neudert, L.-M., Prakash, N., & Vosloo, S. (2021). *Digital Misinformation / Disinformation and Children*. UNICEF Rapid Analysis Report.
- HRW (Human Rights Watch). (2023). *Meta’s Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook*. www.hrw.org/report/2023/12/21/metas-broken-promises/systemic-censorship-palestine-content-instagram-and-facebook
- Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to online disinformation: A framework for cross-national comparative research. *The International Journal of Press/Politics*, 25(3), 493–516. <https://doi.org/10.1177/1940161219900126>
- Humprecht, E., Esser, F., Aelst, P. V., Staender, A., & Morosoli, S. (2023). The sharing of disinformation in cross-national comparison: Analyzing patterns of resilience. *Information, Communication & Society*, 26(7), 1342–1362. <https://doi.org/10.1080/1369118X.2021.2006744>
- Hung, T.-C., & Hung, T.-W. (2022). How China’s cognitive warfare works: A frontline perspective of Taiwan’s anti-disinformation wars. *Journal of Global Security Studies*, 7(4), 1–18. <https://doi.org/10.1093/jogss/ogac016>
- Hunsaker, A., & Hargittai, E. (2018). A review of Internet use among older adults. *New Media & Society*, 20(10), 3937–3954. <https://doi.org/10.1177/1461444818787348>
- Huszár, F., Ktena, S. I., O’Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on Twitter. *PNAS*, 119(1), 1–6. www.pnas.org/doi/epdf/10.1073/pnas.2025334119
- Hyzen, A. (2023). Propaganda and the Web 3.0: Truth and ideology in the digital age. *Nordic Journal of Media Studies*, 5(1), 49–67. <https://doi.org/10.2478/njms-2023-0004>
- Iamnitchi, A., Hall, L. O., Horawalavithana, S., Mubang, F., Ng, K. W., & Skvoretz, J. (2023). Modeling information diffusion in social media: Data-driven observations. *Frontiers in Big Data*, 6, 1–19. <https://doi.org/10.3389/fdata.2023.1135191>
- IEA (2024). Data centres and data transmission networks. IEA (International Energy Agency). www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks
- IEEE (Institute of Electrical and Electronics Engineers). (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*. Version 2. <https://ieeexplore.ieee.org/document/9398613>
- Ihlebæk, K. A., & Sundet, V. S. (2023). Global platforms and asymmetrical power: Industry dynamics and opportunities for policy change. *New Media & Society*, 25(8), 2183–2200. <https://doi.org/10.1177/14614448211029662>
- Ihlebæk, K. A., Figenschou, T. U., Eldridge, S. A., Frischlich, L., Cushion, S., & Holt, K. (2022). Understanding alternative news media and its contribution to diversity. *Digital Journalism*, 10(8), 1267–1282. www.tandfonline.com/doi/abs/10.1080/21670811.2022.2134165

- Iliadis, A., & Acker, A. (2022). The seer and the seen: Surveying Palantir's surveillance platform. *The Information Society*, 38(5), 334–363. www.tandfonline.com/doi/epdf/10.1080/01972243.2022.2100851?needAccess=true
- Innerarity, D. (2024). *Artificial Intelligence and Democracy*. UNESCO Regional Office, Montevideo and CLACSO (Latin American Council of Social Sciences). https://cadmus.eui.eu/bitstream/handle/1814/77333/AI_democracy_2024.pdf?sequence=1
- Innes, H., Dawson, A., & Innes, M. (2023). OSINT vs disinformation: The information threats 'arms race'. Centre for Research and Evidence on Security Threats, 10 October. <https://crestresearch.ac.uk/comment/osint-vs-disinformation-the-information-threats-arms-race>
- International IDEA (Institute for Democracy and Electoral Assistance). (2023). *The Global State of Democracy 2023: The New Checks and Balances*. www.idea.int/gsod/2023
- Introne, J., McKernan, B., Corsbie-Massay, C. L. P., Rohlinger, D., & Tripodi, F. B. (2024). Healthier information ecosystems: A definition and agenda. *Journal of the Association for Information Science and Technology*, 75(10), 1025–1040. <https://doi.org/10.1002/asi.24949>
- IPIE (International Panel on the Information Environment). (2023a). *Countermeasures for Mitigating Digital Misinformation: A Systemic Review*. www.ipie.info/reports-and-publications/reports-and-publications
- IPIE. (2023b). *Platform Responses to Misinformation: A Meta-Analysis of Data*. www.ipie.info/reports-and-publications/reports-and-publications
- IPIE. (2023c). *Trends in the Global Information Environment*. www.ipie.info/reports-and-publications/reports-and-publications
- Ipsos, & UNESCO. (2023). *Survey on the Impact of Online Disinformation and Hate Speech*. www.unesco.org/sites/default/files/medias/fichiers/2023/11/unesco_ipsos_survey_1.pdf
- Ipsos. (2023). *Global Views on AI 2023: How People Across the World Feel about Artificial Intelligence and Expect It Will Impact Their Life: A 31 Country Global Advisory Study*.
- Isin, E., & Ruppert, E. (2020). *Being Digital Citizens*. Rowman & Littlefield.
- IT for Change. (2023a). *Feminist Perspectives on Social Media Governance*. <https://itforchange.net/feminist-perspectives-on-social-media-governance-O>
- IT for Change. (2023b). *IT for Change's Submission to the Forum on Information and Democracy's Working Group on AI and its Implications for the Information & Communication Space*. <https://itforchange.net/it-for-change%E2%80%99s-submission-to-fid-working-group-on-ai-and-its-implications-for-information>
- ITU (International Telecommunication Union). (2024a). *The ICT Development Index 2024: Measuring Digital Development*. www.itu.int/dms_pub/itu-d/opb/ind/D-IND-ICT_MDD-2024-3-PDF-E.pdf
- ITU. (2024b). Statistics ITU. www.itu.int/en/ITU-D/Statistics/pages/stat/default.aspx
- Ivanova, I. (2022). These formerly banned Twitter accounts have been reinstated since Elon Musk took over. CBS News, 21 November. www.cbsnews.com/news/twitter-accounts-reinstated-elon-musk-donald-trump-kanye-ye-jordan-peterson-kathy-griffin-andrew-tate
- Jakob, N. (2012). The tendency to trust as individual predisposition – Exploring the associations between interpersonal trust, trust in the media and trust in institutions. *Communications*, 37(1), 99–120. <https://doi.org/10.1515/commun-2012-0005>
- Jagannatha, A., Rawat, B. P. S., & Yu, H. (2021). Membership inference attack susceptibility of clinical language models. *ArXiv*. <https://doi.org/10.48550/arXiv.2104.08305>
- Jain, A. (2023). We wrote to the Election Commissioner of India on the use of Facial Recognition Technology for voter verification in Karnataka. Internet Freedom Foundation, 13 May. <https://internetfreedom.in/we-wrote-to-the-election-commissioner-of-india-on-the-use-of-facial-recognition-technology-for-voter-verification>
- Jakobsson, P., & Stiernstedt, F. (2023). Trust and the media: Arguments for the (ir)relevance of a concept. *Journalism Studies*, 24(4), 479–495. <https://doi.org/10.1080/1461670X.2023.2169191>
- James, D. (2022). Rigging the rules: How big tech uses stealth 'trade' agreements and how we can stop them. State of Big Tech. <https://projects.itforchange.net/state-of-big-tech/rigging-the-rules-how-big-tech-uses-stealth-trade-agreements-to-undermine-and-prevent-digitalization-in-public-interest-and-how-we-can-stop-them>
- Jamieson, K. H., & Albarracín, D. (2020). The relation between media consumption and misinformation at the outset of the SARS-CoV-2 pandemic in the US. *The Harvard Kennedy School (HKS) Misinformation Review*, 1(2), 1–23. <http://nrs.harvard.edu/urn-3:HUL.InstRepos:42661740>
- Jamieson, K. H., & Cappella, J. N. (2008). *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford University Press.
- Jandrić, P. (2019). The postdigital challenge of critical media literacy. *The International Journal of Critical Media Literacy*, 1(1), 26–37. <https://doi.org/10.1163/25900110-00101002>
- Jasanoff, S. (2015). Future Imperfect: Science, Technology, and the Imaginations of Modernity. In S. Jasanoff & S.-H. Kim (Eds), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power* (pp. 1–34). The University of Chicago Press.

- Jauniaux, L., & Lebourges, M. (2019). Zero rating and end-users' freedom of choice: An economic analysis. *Digital Policy, Regulation and Governance*, 21(2), 115–128. <https://doi.org/10.1108/DPRG-06-2018-0030>
- Jayasinghe, G., Bandaranayake, R., Galpaya, H., Ababakirov, A., & De Silva, R. (2022). Asia. In H. Wasserman (Ed.), *Meeting the Challenges of Information Disorder in the Global South* (pp. 190–245). IDRC Research Results.
- Jeong, S.-H., Cho, H., & Hwang, Y. (2012). Media literacy interventions: A meta-analytic review. *Journal of Communication*, 62(3), 454–472. <https://doi.org/10.1111/j.1460-2466.2012.01643.x>
- Jeppesen, S. (2016). Understanding alternative media power: Mapping content & practice to theory, ideology, and political action. *Democratic Communiqué*, 27(1), 54–77. <https://journals.flvc.org/demcom/article/view/88322>
- Jeppesen, S., Hoechsmann, M., Ulthiin, i. h., VanDyke, D., & McKee, M. (2022). *The Capitol Riots: Digital Media, Disinformation, and Democracy Under Attack*. Routledge.
- Jhaver, S., & Zhang, A. X. (2023). Do users want platform moderation or individual control? Examining the role of third-person effects and free speech support in shaping moderation preferences. *New Media & Society, Online First*, 1–21. <https://doi.org/10.1177/14614448231217993>
- Jiang, B., Tan, Z., Nirmal, A., & Liu, H. (2023). Disinformation detection: An evolving challenge in the age of LLMs. *ArXiv*. <https://doi.org/10.48550/arXiv.2309.15847>
- Jin, Y., & Austin, L. (Eds) (2022). *Social Media and Crisis Communication*. Routledge.
- Jindal, T., & Nigam, A. (2020). *Data Stewardship for Non-Personal Data in India: A Position Paper on Data Trusts*. Vidhi Centre for Legal Policy, Indraprastha Institute of Information Technology, Cyril Amarch and Mangaldas and Omidyar Network India.
- Johnson, G. M. (2023). Are algorithms value-free? Feminist theoretical virtues in machine learning. *Journal of Moral Philosophy*, 21(1–2), 27–61. <https://doi.org/10.1163/17455243-20234372>
- Jones-Jang, S. M., & Chung, M. (2024). Can we blame social media for polarization? Counter-evidence against filter bubble claims during the COVID-19 pandemic. *New Media & Society*, 26(6), 3370–3389. <https://doi.org/10.1177/14614448221099591>
- Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist*, 65(2), 371–388. <https://doi.org/10.1177/0002764219869406>
- Jones, M. O. (2024). News analysis all tweets mentioning both 'Muslim' & 'UK'. X (formerly Twitter), 9 August. <https://x.com/marcowenjones/status/1821881958537871568>
- Jordan, S. (2017). Evaluating zero-rating and associated throttling practices under the Open Internet Order. *Journal of Information Policy*, 7, 450–507. <https://doi.org/10.5325/jinfopoli.7.2017.0450>
- Jørgensen, R. F. (2017). What platforms mean when they talk about human rights. *Policy & Internet*, 9(3), 280–296. <https://doi.org/10.1002/poi3.152>
- Joris, G., Grove, F. D., Van Damme, K., & De Marez, L. (2021). Appreciating news algorithms: Examining audiences' perceptions to different news selection mechanisms. *Digital Journalism*, 9(5), 589–618. <https://doi.org/10.1080/21670811.2021.1912626>
- Jung, E. H., & Sundar, S. S. (2016). Senior citizens on Facebook: How do they interact and why? *Computers in Human Behavior*, 61(2016), 27–35. <https://doi.org/10.1016/j.chb.2016.02.080>
- Jungherr, A., & Schroeder, R. (2021). Disinformation and the structural transformations of the public arena: Addressing the actual challenges to democracy. *Social Media + Society*, 7(1), 1–13. <https://journals.sagepub.com/doi/pdf/10.1177/2056305121988928>
- Jungherr, A., & Schroeder, R. (2023). Artificial intelligence and the public arena. *Communication Theory*, 33(2–3), 164–173. <https://academic.oup.com/ct/article/33/2-3/164/7202294>
- Just, N. (2018). Governing online platforms: Competition policy in times of platformization. *Telecommunications Policy*, 42(5), 386–394. <https://doi.org/10.1016/j.telpol.2018.02.006>
- Just, N. (2022). Which is to be master? Competition law or regulation in platform markets. *International Journal of Communication*, 16(2022), 504–524. <https://ijoc.org/index.php/ijoc/article/view/13095/3657>
- Justice K.S. Puttaswamy (Retd.) & Anr. vs. Union of India & Ors (Judge Bench Civil Writ Petition, Case (2017) 10SCC 1, AIR 2017 SC 4161 2017). <https://privacylibrary.ccnlud.org/case/justice-ks-puttaswamy-ors-vs-union-of-india-ors#:~:text=Case%20Brief&text=The%20Nine%20Judge%20Bench%20in,of%20dignity%2C%20autonomy%20and%20liberty.>
- Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A Flaw in Human Judgment*. William Collins.
- Kakkar, J. M. (2023). Tackling Misinformation in Emerging Economies and the Global South: Exploring Approaches for the Indian Context. In F. Fukuyama & M. Schaake (Eds), *Digital Technologies in Emerging Countries* (pp. 48–62). Cyber Policy Centre, Freeman Spogli Institute and Stanford Law School, Stanford University.
- Kalathil, S., & Boas, T. C. (2003). *Open Networks, Closed Regimes: The Impact of the Internet on Authoritarian Rule*. Carnegie Endowment for International Peace.
- Kamp, C. (2023). *Joint Statement of the informal OSCE Group of Friends on Safety of Journalists*. US Mission to the OSCE. <https://osce.usmission.gov/joint-statement-of-the-informal-osce-group-of-friends-on-safety-of-journalists>
- Kantar Media. (2016). *Brand and Trust in a Fragmented News Environment*. Reuters Institute for the Study of Journalism.

- Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society*, 23(5), 1301–1326. <https://doi.org/10.1177/1461444820959296>
- Karlsen, R., Beyer, A., & Steen-Johnsen, K. (2020). Do high-choice media environments facilitate news avoidance? A longitudinal study 1997–2016. *Journal of Broadcasting & Electronic Media*, 64(5), 794–814. <https://doi.org/10.1080/08838151.2020.1835428>
- Karpf, D. (2020). How Digital Disinformation Turned Dangerous. In S. Livingston & W. L. Bennett (Eds), *The Disinformation Age* (pp. 153–168). Cambridge University Press.
- Kaskazi, A., & Kitzie, V. (2023). Engagement at the margins: Investigating how marginalized teens use digital media for political participation. *New Media & Society*, 25(1), 72–94. <https://doi.org/10.1177/14614448211009460>
- Katsirea, I. (2018). ‘Fake news’: Reconsidering the value of untruthful expression in the face of regulatory uncertainty. *Journal of Media Law*, 10(2), 159–188. <https://doi.org/10.1080/17577632.2019.1573569>
- Katz, E. (1957). The two-step flow of communication: An up-to-date report on an hypothesis. *Public Opinion Quarterly*, 21(1), 61–78. <https://doi.org/10.1086/266687>
- Katzenbach, C. (2021). ‘AI will fix this’ – The technical, discursive, and political turn to AI in governing communication. *Big Data & Society*, 8(2), 1–8. <https://doi.org/10.1177/20539517211046182>
- Kaye, D. (2015). *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*. UN Human Rights Council. May.
- Kaye, D. (2018). *Mandate of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression: Letter by David Kay*. OHCHR Letter OL-ITA-1-2018.
- Keller, T. R., & Klinger, U. (2019). Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Political Communication*, 36(1), 171–189. <https://doi.org/10.1080/10584609.2018.1526238>
- Kennedy, I., Wack, M., Beers, A., Schafer, J. S., et al. (2022). Repeat spreaders and election delegitimization: A comprehensive dataset of misinformation Tweets from the 2020 US election. *Journal of Quantitative Description: Digital Media*, 2(2022), 1–49. <https://doi.org/10.51685/jqd.2022.013>
- Kennis, A. (2024). Elite News Content (Still) Matters Now More than Ever Before: Social Movements, Social Media, Elite News Media, and the Media Dependence Model. In J. Pedro-Carañana, R. Gómez, T. F. Corrigan, & F. R. Sierra Caballero (Eds), *Political Economy of Media and Communication: Methodological Approaches* (pp. 224–244). Routledge.
- Kerber, W. (2020). *From (Horizontal and Sectoral) Data Access Solutions towards Data Governance Systems*. Joint Discussion Paper No. 40–2020. Series in Economics. Universities of Aachen, Gießen, Göttingen Kassel, Marburg, Siegen.
- Keremoğlu, E., & Weidmann, N. B. (2020). How dictators control the internet: A review essay. *Comparative Political Studies*, 53(10–11), 1690–1703. <https://doi.org/10.1177/0010414020912278>
- Kertysova, K. (2018). Artificial intelligence and disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered. *Security and Human Rights*, 29(1–4), 55–81. <https://doi.org/10.1163/18750230-02901005>
- Kettemann, M. C. (2022). *The Normative Order of the Internet. A Theory of Rule and Regulation Online*. Oxford University Press.
- Kettemann, M. C., & Lachmayer, K. (Eds) (2021). *Pandemocracy in Europe: Power, Parliaments and People in Times of COVID-19*. Bloomsbury.
- Kettemann, M. C., & Schulz, W. (2023). *Platform://Democracy – Perspectives on Platform Power, Public Values and the Potential of Social Media Councils*. Verlag Hans-Bredow-Institut.
- Kim, K., & Moon, S.-I. (2021). When algorithmic transparency failed: Controversies over algorithm-driven content curation in the South Korean digital environment. *American Behavioral Scientist*, 65(6), 847–862. <https://doi.org/10.1177/0002764221989783>
- Kim, S. J. (2016). A repertoire approach to cross-platform media use behavior. *New Media & Society*, 18(3), 353–372. <https://doi.org/10.1177/1461444814543162>
- Kinzelbach, K., Lindberg, S. I., & Lott, L. (2024). *Academic Freedom Index Update 2024*. Friedrich-Alexander-Universität and V-Dem Institute.
- Kira, B., Sinha, V., & Srinivasan, S. (2021). Regulating digital ecosystems: Bridging the gap between competition policy and data protection. *Industrial and Corporate Change*, 30(5), 1337–1360. <https://academic.oup.com/icc/article/30/5/1337/6356942>
- Klapper, J. T. (1960). *The Effects of Mass Communication*. Free Press.
- Klein, N. (2008). *The Shock Doctrine: The Rise of Disaster Capitalism*. Picador.
- Klein, N. (2020). Naomi Klein: How big tech plans to profit from the pandemic. *The Guardian*, 13 May. www.theguardian.com/news/2020/may/13/naomi-klein-how-big-tech-plans-to-profit-from-coronavirus-pandemic
- Klonick, K. (2017). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 131(6), 1598–1670. https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf
- Knight, W. (2024). OpenAI employees warn of a culture of risk and retaliation. *Wired*, 4 June. www.wired.com/story/openai-right-to-warn-open-letter-ai-risk

- Knuutila, A., Neudert, L.-M., & Howard, P. N. (2022). Who is afraid of fake news? Modeling risk perceptions of misinformation in 142 countries. *Harvard Kennedy School (HKS) Misinformation Review*, 3(3), 1–13. <https://doi.org/10.37016/mr-2020-97>
- Koene, A., Smith, A. L., Egawa, T., Mandalh, S., & Hatada, Y. (2018). IEEE P70xx, establishing standards for ethical technology. *Proceedings of KDD*. <https://doi.org/10.1145/nnnnnnnn.nnnnnnn>
- Kohring, M., & Matthes, J. (2007). Trust in news media: Development and validation of a multidimensional scale. *Communication Research*, 34(2), 231–252. <https://doi.org/10.1177/0093650206298071>
- Kokas, A. (2022). Becoming a Cyber Sovereign: China's Politics of Data Governance. In A. Kokas, *Trafficking Data: How China Is Winning the Battle for Digital Sovereignty* (pp. 51–C53.P72). Oxford Academic.
- Kokshagina, O., Reinecke, P. C., & Karanasios, S. (2023). To regulate or not to regulate: Unravelling institutional tussles around the regulation of algorithmic control of digital platforms. *Journal of Information Technology*, 38(2), 160–179. <https://doi.org/10.1177/0268396222114408>
- Kop, M. (2020). *Shaping the Law of AI: Transatlantic Perspectives*. Stanford Law School. <https://law.stanford.edu/publications/no-65-shaping-the-law-of-ai-transatlantic-perspectives>
- Kostygina, G., Kim, Y., Seeskin, Z., LeClere, F., & Emery, S. (2023). Disclosure standards for social media and generative artificial intelligence research: Toward transparency and replicability. *Social Media + Society*, 9(4), 1–12. <https://doi.org/10.1177/20563051231216947>
- Kothari, A., & Cruikshank, S. A. (2022). Artificial intelligence and journalism: An agenda for journalism research in Africa. *African Journalism Studies*, 43(1), 17–33. <https://doi.org/10.1080/23743670.2021.1999840>
- Kozyreva, A., Lorenz-Spreen, P., Herzog, S. M., Ecker, U. K. H., Lewandowsky, S., & Hertwig, R., et al. (2024). Toolbox of individual-level interventions against online misinformation. *Nature Human Behavior*, 8(6), 1044–1052. <https://www.nature.com/articles/s41562-024-01881-0>
- Krämer, J., & Peitz, M. (2018). A fresh look at zero-rating. *Telecommunications Policy*, 42(7), 501–513. <https://doi.org/10.1016/j.telpol.2018.06.005>
- Kreiss, D., & McGregor, S. C. (2023). A review and provocation: On polarization and platforms. *New Media & Society*, 26(1), 556–579. <https://doi.org/10.1177/14614448231161880>
- Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, 9(1), 104–117. <https://doi.org/10.1017/XPS.2020.37>
- Kristensen, L. M., & Hartley, J. M. (2023). The infrastructure of news: Negotiating infrastructural capture and autonomy in data-driven news distribution. *Media and Communication*, 11(2), 307–318. <https://doi.org/10.17645/mac.v11i2.6388>
- Kroet, C. (2024). Meta urged to answer EU questions on CrowdTangle. *Euronews*, 6 August. www.euronews.com/next/2024/08/16/meta-urged-to-answer-eu-questions-on-crowdtangle
- Krouwel, A., Kutyski, Y., van Prooijen, J.-W., Martinsson, J., & Markstedt, E. (2017). Does extreme political ideology predict conspiracy beliefs, economic evaluations and political trust? Evidence From Sweden. *Journal of Social and Political Psychology*, 5(2), 435–462. <https://doi.org/10.5964/jspp.v5i2.745>
- Książak, P., & Wojtczak, S. (2023). *Toward a Conceptual Network for the Private Law of Artificial Intelligence* (Vol. 51). Springer International Publishing.
- Kulesz, M. (2018). *Culture, Platforms and Machines: The Impact of Artificial Intelligence on the Diversity of Cultural Expressions*. UNESCO Report to Intergovernmental Committee for the Protection and Promotion of the Diversity of Cultural Expressions, DCE/18/12.IGC/INF.4.
- Kumar, A., & Thussu, D. (2023). Media, digital sovereignty and geopolitics: The case of the TikTok ban in India. *Media, Culture & Society*, 45(8), 1583–1599. <https://doi.org/10.1177/01634437231174351>
- Kumar, R., Desai, V. S., & Koshy, N. S. (2023). *Creating Sustainable Data Cooperatives in the Global South: Frameworks for Institutional Support: Research Brief*. IT for Change, India.
- Kuznetsova, E., & Makhortykh, M. (2023). Blame It on the algorithm? Russian government-sponsored media and algorithmic curation of political information on Facebook. *International Journal of Communication*, 17(2023), 971–992. <https://ijoc.org/index.php/ijoc/article/view/18687>
- Kyriakidou, M., Cushion, S., Hughes, C., & Morani, M. (2023). Questioning fact-checking in the fight against disinformation: An audience perspective. *Journalism Practice*, 17(10), 2123–2139. <https://doi.org/10.1080/17512786.2022.2097118>
- la Cour, C. (2020). Theorising digital disinformation in international relations. *International Politics*, 57(4), 704–723. <https://doi.org/10.1057/s41311-020-00215-x>
- Labio-Bernal, A., & Manzano-Zambruno, L. (2023). 8M demonstrations, the Spanish far right and the pandemic in a hybrid media system. *Media and Communication*, 11(1), 114–124. <https://idus.us.es/handle/11441/142909>
- Labio-Bernal, A., & Romero-Domínguez, L. R. (2022). *Periodismo SOS – Del filantropocapitalismo al poder que mata*. Ediciones Pirámide.
- LaForge, G., & Gruver, P. (2023). Governing the Digital Future. *Planetary Politics*, October. <https://core.ac.uk/download/588303818.pdf>
- Lamdan, S. (2022). *Data Cartels: The Companies that Control and Monopolize Our Information*. Stanford University Press.
- Landau, M. (2021). The Astounding Growth of 'Big Tech' and the Lack of Enforcement of the Intellectual Property, Antitrust and Contract Laws. In K. B. Tripp (Ed.), *Entertainment, Publishing and the Arts Handbook*. Clark Boardman Callaghan.

- Langer, A. I., & Gruber, J. B. (2021). Political agenda setting in the hybrid media system: Why legacy media still matter a great deal. *The International Journal of Press/Politics*, 26(2), 313–340. <https://doi.org/10.1177/1940161220925023>
- Lanier, J., & Weyl, E. G. (2018). A blueprint for a better digital society. *Harvard Business Review*, 26 September, 2–19. <https://hbr.org/2018/09/a-blueprint-for-a-better-digital-society>
- Lanzing, M. (2023). Traveling technology and perverted logics: Conceptualizing Palantir’s expansion into health as sphere transgression. *Information, Communication and Society*, 1–17. <https://doi.org/10.1080/1369118x.2023.2279557>
- Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLOS ONE*, 10(6), 1–15. <https://doi.org/10.1371/journal.pone.0127502>
- Larsen, B. C., & Küspert, S. (2024). Regulating general-purpose AI: Areas of convergence and divergence across the EU and the US. Brookings Institution Research, 21 May. www.brookings.edu/articles/regulating-general-purpose-ai-areas-of-convergence-and-divergence-across-the-eu-and-the-us
- Lasswell, H. D. (1971). *Propaganda Technique in the World War*. MIT Press.
- Lauer, L., & Graves, L. (2024). How to grow a transnational field: A network analysis of the global fact-checking movement. *New Media & Society, Online First*, 1–19. <https://doi.org/10.1177/14614448241227856>
- Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3(2022), 1–15. <https://doi.org/10.1016/j.caeai.2022.100101>
- Lawder, D. (2023, 2023/11/08/). US suspends Indo-Pacific talks on key aspects of digital trade –lawmakers. Reuters, 8 November. www.reuters.com/business/finance/us-suspends-indo-pacific-talks-key-aspects-digital-trade-lawmakers-2023-11-08
- Lawrence, S. E.-L. (2022). ‘The Majority World – what’s in a phrase?’ Philanthropy Australia, Blog, 4 November. www.philanthropy.org.au/news-and-stories/the-majority-world-whats-in-a-phrase
- Lecheler, S., & de Vreese, C. H. (2017). News media, knowledge, and political interest: Evidence of a dual role from a field experiment: News media, knowledge, and political interest. *Journal of Communication*, 67(4), 545–564. <https://doi.org/10.1111/jcom.12314>
- Ledford, H. (2019). Millions of black people affected by racial bias in health-care algorithms. *Nature*, 574(7780), 608–609. <https://doi.org/10.1038/d41586-019-03228-6>
- Lee, D., & Shin, D.-H. (2016). The effects of network neutrality on the incentive to discriminate, invest, and innovate: a literature review. *info*, 18(3), 42–57. <https://doi.org/10.1108/info-12-2015-0053>
- Lee, J. (2024). CCPA/CPRA: Consumers bear the burden as companies bear the crown. *Hastings International & Comparative Law Review*, 47(2), 129–152. https://repository.uclawsf.edu/hastings_international_comparative_law_review/vol47/iss2/5
- Lee, S., & Valenzuela, S. (2024). A self-righteous, not a virtuous, circle: Proposing a new framework for studying media effects on knowledge and political participation in a social media environment. *Social Media + Society*, 10(2), 1–7. <https://doi.org/10.1177/20563051241257953>
- Lehdonvirta, V. (2022). *Cloud Empires: How Digital Platforms Are Overtaking the State and How We Can Regain Control*. MIT Press.
- Lehedé, S. (2022). Territories of data: Ontological divergences in the growth of data infrastructure. *Tapuya: Latin American Science, Technology and Society*, 5(1), 1–18. <https://doi.org/10.1080/25729861.2022.2035936>
- Lemke, T., & Habegger, M. W. (2022). Foreign interference and social media networks: A relational approach to studying contemporary Russian disinformation. *Journal of Global Security Studies*, 7(2), 1–25. <https://doi.org/10.1093/jogss/ogac004>
- Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, 31(4), 611–627. <https://doi.org/10.1007/s13347-017-0279-x>
- Lesh, M. (2023). *Breaking the News? Should Digital Platforms Be Required to Fund News Publishers?* IEA Discussion paper No. 119. Institute of Economic Affairs.
- Lewandowsky, S., Armaos, K., Bruns, H., Schmid, P., et al. (2022). When science becomes embroiled in conflict: Recognizing the public’s need for debate while combating conspiracies and misinformation. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 26–40. <https://doi.org/10.1177/00027162221084663>
- Lewkowiz, G., & Sarf, R. (2024). Taking technical standardization of fundamental rights seriously for trustworthy artificial intelligence. *La Revue des Juristes de Sciences Po*, 25, 42–46. <https://centrereliman.be/content/uploads/2024/03/GL-RS-AI-Technical-standards-and-fundamental-rights.pdf>
- LexisNexis. (2024). Sources of local and international news and business information. www.lexisnexis.ca/en-ca/products/nexis.page
- Li, W. Y. (2023). Regulatory capture’s third face of power. *Socio-Economic Review*, 21(2), 1217–1245. <https://academic.oup.com/ser/article/21/2/1217/7030814?login=false>
- Lim, G., & Bradshaw, S. (2023). *Chilling Legislation: Tracking the Impact of ‘Fake News’ Laws on Press Freedom Internationally*. Center for International Media Assistance.
- Lin, B., & Lewis, S. C. (2022). The one thing journalistic AI just might do for democracy. *Digital Journalism*, 10(10), 1627–1649. <https://doi.org/10.1080/21670811.2022.2084131>

- Linden, T., Khandelwal, R., Harkous, H., & Fawaz, K. (2020). The privacy policy landscape after the GDPR. *ArXiv*. <https://doi.org/10.48550/arXiv.1809.083696>
- Linehan, M., Byers, C., Brooks, N. N., & Freeman, L. (2024). *Responsible Generative AI: An Examination of Ongoing Efforts to Tame This Powerful Technology*. OMG and IBM.
- Linvill, D. L., & Warren, P. L. (2019). *Touched by the Trolls: How and Why a Coordinated Information Operation Interacts with Outsiders*. www.semanticscholar.org/paper/Touched-by-the-Trolls-%3A-How-and-Why-a-Coordinated-Linvill-Warren/deba83d06283496135f5284e86201039ead352d3
- Linvill, D. L., & Warren, P. L. (2020). Troll factories: Manufacturing specialized disinformation on Twitter. *Political Communication*, 37(4), 447–467. <https://doi.org/10.1080/10584609.2020.1718257>
- Liu, Y., Zhang, K., Li, Y., Yan, Z., et al. (2024). Sora: A review on background, technology, limitations, and opportunities of large vision models. *ArXiv*. <https://doi.org/10.48550/arXiv.2402.17177>
- Liveriero, F. (2020). Epistemic injustice in the political domain: Powerless citizens and institutional reform. *Ethical Theory and Moral Practice*, 23(5), 797–813. <https://doi.org/10.1007/s10677-020-10097-w>
- Livingstone, S. (1996). On the Continuing Problems of Media Effects Research. In J. Curran & M. Gurevitch (Eds), *Mass Media and Society* (pp. 305–324). Edward Arnold.
- Livingstone, S. (1998). Relationships between Media and Audiences: Prospects for Audience Reception Studies. In T. Liebes & J. Curran (Eds), *Media, Ritual and Identity: Essays in Honor of Elihu Katz* (pp. 237–255). Routledge.
- Livingstone, S. (2008). Engaging with media – a matter of literacy? *Communication, Culture & Critique*, 1(1), 51–62. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1753-9137.2007.00006.x>
- Livingstone, S., & Blum-Ross, A. (2020). *Parenting for a Digital Future: How Hopes and Fears about Technology Shape Children's Lives*. Oxford University Press.
- Livingstone, S. (2023). What's the best we can expect of media literacy? From protectionism to human rights and flourishing. Media Literacy Blog. <https://blogs.lse.ac.uk/medialse/2023/12/13/whats-the-best-we-can-expect-of-media-literacy-from-protectionism-to-human-rights-and-flourishing>
- Livingstone, S., Mascheroni, G., & Stoilova, M. (2023a). The outcomes of gaining digital skills for young people's lives and wellbeing: A systematic evidence review. *New Media & Society*, 25(5), 1176–1202. <https://doi.org/10.1177/14614448211043189>
- Livingstone, S., Ólafsson, K., & Pothong, K. (2023b). Digital play on children's terms: A child rights approach to designing digital experiences. *New Media & Society, Online First*, 1–21. <https://doi.org/10.1177/14614448231196579>
- Livingstone, S., Cantwell, N., Özkul, D., Shekhawat, G., & Kidron, B. (2024). *The Best Interests of the Child in the Digital Environment*. Digital Futures for Children Centre, London School of Economics and Political Science and 5Rights Foundation.
- Livio, O., & Cohen, J. (2018). 'Fool me once, shame on you': Direct personal experience and media trust. *Journalism*, 19(5), 684–698. <https://doi.org/10.1177/1464884916671331>
- Llansó, E. J., van Hoboken, J., & Harambam, J. (2020). *Artificial Intelligence, Content Moderation, and Freedom of Expression*. TransAtlantic Working Group on Content Moderation Online and Freedom of Expression, Annenberg Public Policy Center and Institute for Information Law, University of Amsterdam.
- Local Progress. (2024). Creating Traffic Safety: A Policy Memo for Local Elected Leaders. 9 January. <https://localprogress.org/resource/creating-traffic-safety-a-policy-memo-for-local-elected-leaders>
- Lockie, S. (2017). Post-truth politics and the social sciences. *Environmental Sociology*, 3(1), 1–5. <https://doi.org/10.1080/23251042.2016.1273444>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *CHI'20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3313831.3376727>
- Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., et al. (2024). Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *ArXiv*. <https://doi.org/10.1016/j.inffus.2024.102301>
- Loos, E., & Ivan, L. (2023). Using Media Literacy to Fight Digital Fake News in Later Life: A Mission Impossible? In Q. Gao & J. Zhou (Eds), *Human Aspects of IT for the Aged Population. HCII 2023. Lecture Notes in Computer Science*, vol 14042. Springer. https://doi.org/10.1007/978-3-031-34866-2_18
- LSE Truth, Trust & Technology Commission. (2018). *Tackling the Information Crisis: A Policy Framework for Media System Resilience*. LSE Commission on Truth, Trust and Technology.
- Luchs, I. (2023). AI for all? Challenging the democratization of machine learning. *Minor Tech: A Peer-Reviewed Journal About AI*, 12(1), 135–145. <https://doi.org/10.7146/aprja.v12i1.140445>
- Lunga, C. M., & Mthembu, M. V. (2019). Investigating the source and strategies adopted by mainstream media in combating fake news in the Kingdom of Eswatini. *African Journalism Studies*, 40(4), 96–111. <https://doi.org/10.1080/23743670.2019.1664606>
- Mabweazara, H. M., & Pearson, B. (Eds) (2024). *Media Capture in Africa and Latin America: Power and Resistance*. Springer International.
- Mabweazara, H. M., Muneri, C. T., & Ndlovu, F. (2020). News 'media capture', relations of patronage and clientelist practices in sub-Saharan Africa: An interpretive qualitative analysis. *Journalism Studies*, 21(15), 2154–2175. <https://doi.org/10.1080/1461670X.2020.1816489>

- MacCarthy, M. (2023). *Regulating Digital Industries: How Public Oversight Can Encourage Competition, Protect Privacy, and Ensure Free Speech*. Brookings Institution Press.
- Macdonald, S., Correia, S. G., & Watkin, A.-L. (2019). Regulating terrorist content on social media: Automation and the rule of law. *International Journal of Law in Context*, 15(2), 183–197. <https://doi.org/10.1017/S1744552319000119>
- Machill, S. A. (2020). *Biased Artificial Intelligence: Algorithmic Fairness and Human Perception of Biased AI*. NOVA Information Management School, Universidade Nova de Lisboa.
- MacKenzie, D., Caliskan, K., & Rommerskirchen, C. (2023). The longest second: Header bidding and the material politics of online advertising. *Economy and Society*, 52(3), 554–578. <https://doi.org/10.1080/03085147.2023.2238463>
- Madan, G. D. (2021). Understanding misinformation in India: The case for a meaningful regulatory approach for social media platforms. *ArXiv*. <https://doi.org/10.48550/arXiv.2207.01508>
- Madrid-Morales, D., & Wasserman, H. (2022). Research methods in comparative disinformation studies. In H. Wasserman & D. Madrid-Morales (Eds.), *Disinformation in the Global South* (pp. 41–57). Wiley Blackwell.
- Madrid-Morales, D., Wasserman, H., Gondwe, G., Ndlovu, K., et al. (2021). Motivations for sharing misinformation: A comparative study in six sub-Saharan African countries. *International Journal of Communication*, 15, 1200–1219. <https://ijoc.org/index.php/ijoc/article/view/14801>
- Magalhães, J. C. (2018). Do algorithms shape character? Considering algorithmic ethical subjectivation. *Social Media + Society*, 4(2), 1–10. <https://doi.org/10.1177/2056305118768301>
- Magalhães, J. C., & Couldry, N. (2021). Giving by taking away: Big tech, data colonialism, and the reconfiguration of social good. *International Journal of Communication*, 15(2021), 343–362. <https://ijoc.org/index.php/ijoc/article/view/15995>
- Mager, A., & Katzenbach, C. (2021). Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified. *New Media & Society*, 23(2), 223–236. <https://doi.org/10.1177/1461444820929321>
- Mahler, T. (2022). Between Risk Management and Proportionality: The Risk-Based Approach in the EU's Artificial Intelligence Act Proposal. In L. Colonna & S. Greenstein (Eds.), *Law in the Era of Artificial Intelligence: Nordic Yearbook of Law and Informatics 2020–2021* (pp. 247–270). The Swedish Law and Informatics Research Institute.
- Mahomed, S., Aitken, M., Atabey, A., Wong, J., & Briggs, M. (2023). *AI, Children's Rights, & Wellbeing: Transnational Frameworks: Mapping 13 Frameworks at the Intersections of Data-Intensive Technologies, Children's Rights and Wellbeing*. Alan Turing Institute UK.
- Majal (2024a). About us. <https://majal.org/about-us>
- Majal (2024b). Numun Fund: Supporting feminist tech infrastructure. <https://majal.org/numun-fund>
- Makan, N. (2023). Sustainable industrial machine learning and artificial intelligence: A framework for environmentally responsible innovation. *Transactions on Recent Developments in Artificial Intelligence and Machine Learning*, 15(15). <https://journals.threms.com/index.php/TRDAIML/article/view/184>
- Makhortykh, M., Urman, A., Münch, F. V., Heldt, A., Dreyer, S., & Kettemann, M. C. (2022). Not all who are bots are evil: A cross-platform analysis of automated agent governance. *New Media & Society*, 24(4), 964–981. <https://doi.org/10.1177/14614448221079035>
- Malaysia Government. (2018). *Anti-Fake News Bill 2018*.
- Mano, W., & Milton, V. C. (Eds) (2021). *Routledge Handbook of African Media and Communication Studies*. Routledge.
- Mansell, R. (2012). *Imagining the Internet: Communication, Innovation and Governance*. Oxford University Press.
- Mansell, R. (2021). European Responses to (US) Digital Platform Dominance. In D. Y. Jin (Ed.), *The Routledge Handbook of Digital Media and Globalization* (pp. 141–149). Routledge.
- Mansell, R., & Nordenstreng, K. (2006). Great media and communications debates – An assessment of the MacBride Report after 25 years. *Information Technologies and International Development*, 3(4), 15–36. <https://itidjournal.org/index.php/itid/article/view/235.html>
- Mansell, R., & Steinmueller, W. E. (2020). *Advanced Introduction to Platform Economics*. Edward Elgar Publishing.
- Marconi, F. (2020). *Newsmakers – Artificial Intelligence and the Future of Journalism*. Columbia University Press.
- Marcucci, S., Alarcón, N. G., Verhulst, S. G., & Wüllhorst, E. (2023). Informing the global data future: Benchmarking data governance frameworks. *Data & Policy*, 5(e30), 1–25. <https://doi.org/10.1017/dap.2023.24>
- Marda, V., & Milan, S. (2018). *Wisdom of the Crowd: Multistakeholder Perspectives on the Fake News Debate*. Internet Policy Observatory, Annenberg School for Communication, University of Pennsylvania and DATACTIVE Ideas Lab, NL.
- Mare, A. (2020). State-ordered internet shutdowns and digital authoritarianism in Zimbabwe. *International Journal of Communication*, 14(2020), 4244–4263. <https://ijoc.org/index.php/ijoc/article/view/11494>
- Mare, A., Mabweazara, H. M., & Moyo, D. (2019). 'Fake news' and cyber-propaganda in sub-Saharan Africa: Recentering the research agenda. *African Journalism Studies*, 40(4), 1–12. <https://doi.org/10.1080/23743670.2020.1788295>
- Marinova, Z., Spangenberg, J., Teyssou, D., Papadopoulos, S., et al. (2020). Weverify: Wider and Enhanced Verification for You Project Overview and Tools. In *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. <https://ieeexplore.ieee.org/document/9106056>
- Marsden, C. T. (2016). Comparative case studies in implementing net neutrality: A critical analysis of zero rating. *SCRIPTed: A Journal of Law, Technology & Society*, 13(1), 1–39. <https://doi.org/10.2966/scrip.130116.1>

- Marsden, C. T., & Brown, I. (2023). App stores, antitrust and their links to net neutrality: A review of the European policy and academic debate leading to the EU Digital Markets Act. *Internet Policy Review*, 12(1), 1–27. <https://policyreview.info/articles/analysis/app-stores-antitrust-net-neutrality-eu-digital-markets-act>
- Marshall, S. (2023). We get past ‘post-platform’. Nieman Lab. www.niemanlab.org/2023/12/we-get-past-post-platform
- Martens, M., De Wolf, R., Berendt, B., & De Marez, L. (2023). Decoding algorithms: Exploring end-users’ mental models of the inner workings of algorithmic news recommenders. *Digital Journalism*, 11(1), 203–225. <https://doi.org/10.1080/21670811.2022.2129402>
- Martin, K. (2022). Google Research: Who Is Responsible for Ethics of AI? In K. Martin (Ed.), *Ethics of Data and Analytics: Concepts and Cases* (pp. 13). Auerbach Publications.
- Martínez-Costa, M.-P., López-Pan, F., Buslón, N., & Salaverría, R. (2023). Nobody-fools-me perception: Influence of age and education on overconfidence about spotting disinformation. *Journalism Practice*, 17(10), 2084–2102. <https://doi.org/10.1080/17512786.2022.2135128>
- Masur, P. K. (2020). How online privacy literacy supports self-data protection and self-determination in the age of information. *Media and Communication*, 8(2), 1–12. www.ssoar.info/ssoar/handle/document/69359
- Masur, P. K., DiFranzo, D., & Bazarova, N. N. (2021). Behavioral contagion on social media: Effects of social norms, design interventions, and critical media literacy on self-disclosure. *PLOS ONE*, 16(7). <https://doi.org/10.1371/journal.pone.0254670>
- Matasick, C., Alfonsi, C., & Bellantoni, A. (2020). *Governance Responses to Disinformation: How Open Government Principles Can Inform Policy Options*. OECD.
- Mateen, H., Tabakovic, H., Holder, P., & Schiffrin, A. (2023). *Paying for News: What Google and Meta Owe Publishers*. Initiative for Policy Dialogue, Columbia University. <http://dx.doi.org/10.2139/ssrn.4704237>
- Mattelart, T. (2023). US Digital Platforms in the Global South: A Critical Review of an Emerging Research Field. In P. Bouquillion, C. Ithurbide, & T. Mattelart (Eds), *Digital Platforms and the Global South: Reconfiguring Power Relations in the Cultural Industries* (pp. 18–36). Routledge.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution that Will Transform How We Live, Work and Think*. John Murray.
- Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On Faithfulness and Factuality in Abstractive Summarization. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1906–1919). Association for Computational Linguistics.
- Mayson, S. G. (2018). Bias in, bias out. *The Yale Law Journal*, 2393, 2218–2300. https://scholarship.law.upenn.edu/faculty_scholarship/2393
- Mazzoli, E. M. (2020). Online content governance: Towards a framework for analysis for prominence and discoverability. *Journal of Digital Media & Policy*, 11, 301–319. https://intellectdiscover.com/content/journals/10.1386/jdmp_00027_1
- Mazzoli, E. M. (2023). *Digital Trust Initiatives: Seeking to Reward Journalistic Ethics Online*. Center for International Media Assistance and National Endowment for Democracy, US.
- Mazzucato, M. (2023). Governing the Economics of the Common Good: From Correcting Market Failures to Shaping Collective Goals. Working Paper 2023/08. *Institute for Innovation and Public Purpose*. www.ucl.ac.uk/bartlett/public-purpose/sites/bartlett_public Purpose/files/mazzucato_m_2023_governing_the_economics_of_the_common_good_from_correcting_market_failures_to_shaping_collective_goals.pdf
- McCombs, M., & Shaw, D. L. (1972). The agenda-setting function of mass media. *The Public Opinion Quarterly*, 36(2), 176–187. <https://academic.oup.com/poq/article/36/2/176/1853310>
- McConvey, J. R. (2024). Police in US cities that ban facial recognition asking others to do it for them. *Biometric Update*, 22 May. www.biometricupdate.com/202405/police-in-us-cities-that-ban-facial-recognition-asking-others-to-do-it-for-them
- McGonagle, T., Bednarski, M., Francese Coutinho, M., & Zimin, A. (2019). *Elections and Media in Digital Times*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000371486>
- McGuigan, L. (2023). *Selling the American People: Advertising, Optimization, and the Origins of Adtech*. MIT Press.
- McIntyre, K., & Sobel, M. (2019). How Rwandan journalists use WhatsApp to advance their profession and collaborate for the good of their country. *Digital Journalism*, 7(6), 705–724. <https://doi.org/10.1080/21670811.2019.1612261>
- McKernan, B., Rossini, P., & Stromer-Galley, J. (2023). Echo chambers, cognitive thinking styles, and mistrust? Examining the roles information sources and information processing play in conspiracist ideation. *International Journal of Communication*, 17(2023), 1102–1125. <https://ijoc.org/index.php/ijoc/article/view/19244>
- McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. McGraw-Hill.
- McQuail, D. (2010). *Mass Communication Theory*. 6th Edition. SAGE Publications.
- Medrado, A., & Verdegem, P. (2024). Participatory action research in critical data studies: Interrogating AI from a South–North approach. *Big Data & Society*, 11(1), 1–14. <https://doi.org/10.1177/20539517241235869>
- Medronho, R. (2024). Far-right seeks to censor science and criminalize researchers studying disinformation in Brazil. *Tech Policy Press*, 14 June. <https://techpolicy.press/farright-seeks-to-censor-science-and-criminalize-researchers-studying-disinformation-in-brazil>

- Meese, J., & Hurcombe, E. (2021). Facebook, news media and platform dependency: The institutional impacts of news distribution on social platforms. *New Media & Society*, 23(8), 2367–2384. <https://doi.org/10.1177/1461444820926472>
- Mehta, S., & Erickson, K. (2022). Can online political targeting be rendered transparent? Prospects for campaign oversight using the Facebook Ad Library. *Internet Policy Review*, 11(1), 1–32. <https://doi.org/10.14763/2022.1.1648>
- MeitY (Ministry of Electronics and Information Technology) India. (2023). *The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 [updated 6 April 2023]*. Government of India.
- Mejias, U. A., & Couldry, N. (2024). *Data Grab: The New Colonialism of Big Tech and How to Fight Back*. The University of Chicago Press.
- Melih, Y. (2022). New geographies of platform capitalism: The case of digital monopolization in Turkey. *Big Data & Society*, 9(2), 1–14. <https://doi.org/10.1177/20539517221124585>
- Mendes, L. S., & Kira, B. (2023). The road to regulation of artificial intelligence: The Brazilian experience. *Internet Policy Review*, 21 December. <https://policyreview.info/articles/news/road-regulation-artificial-intelligence-brazilian-experience/1737>
- Mendoza, R. U., Elemia, C. K. S., Recto, J. M. M., & de Castro, B. A. B. (2023). When fake news infects political networks: Case study of the Tallano gold myth in the Philippines. *Media Asia*, 50(4), 501–527. <https://doi.org/10.1080/01296612.2023.2217607>
- Menn, J., & Nix, N. (2023). Big Tech funds the very people who are supposed to hold it accountable. *The Washington Post*, 6 December. www.washingtonpost.com/technology/2023/12/06/academic-research-meta-google-university-influence/
- Menon, S. (2017). The role of civil society groups in improving access to the DC–CAN (District of Columbia – Community Access Network). *Information Systems Frontiers*, 19(2), 361–375. <https://doi.org/10.1007/s10796-015-9602-1>
- Menon, S. (2021). An institutional analysis of TMP regulation in India. *Review of Policy Research*, 38(3), 300–325. <https://doi.org/10.1111/ropr.12425>
- Menon, S. (2024). The digital navigator programme in the time of COVID-19: A case study on Philadelphia’s programme. *Journal of Digital Media & Policy, Online First*, 1–26. https://doi.org/10.1386/jdmp_00140_1
- Merenda, F. (2021). Reading Arendt to Rethink Truth, Science and Politics in the Era of Fake News. In S. Giusti & E. Piras (Eds), *Democracy and Fake News: Information Manipulation and Post-Truth Politics* (pp. 19–29). Routledge.
- Metz, C. (2023). The ChatGPT king isn’t worried, but he knows you might be. *The New York Times*, 31 March. www.nytimes.com/2023/03/31/technology/sam-altman-open-ai-chatgpt.html
- Metzger, M. J., Hartsell, E. H., & Flanagin, A. J. (2020). Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news. *Communication Research*, 47(1), 3–28. <https://doi.org/10.1177/0093650215613136>
- Michael, A. (2023). Artificial intelligence, democracy and elections. European Parliament Briefing. [www.europarl.europa.eu/RegData/etudes/BRIE/2023/751478/EPRS_BRI\(2023\)751478_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/751478/EPRS_BRI(2023)751478_EN.pdf)
- Michalis, M., & D’Arma, A. (2024). Public Service Media: From Epistemic Rights to Justice. In M. A. Horowitz, H. Nieminen, K. Lehtisaari, & A. D’Arma (Ed.), *Epistemic Rights in the Era of Digital Disruption* (pp. 97–109). Springer International Publishing.
- Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, 7(2), 1–15. <https://doi.org/10.1177/2053951720948087>
- Microsoft. (2023). *Diversity and Inclusion Report 2023: A Decade of Transparency, Commitment and Progress*.
- Miguel, R., & Krack, N. (2023). Platforms’ policies on AI-manipulated and generated misinformation. EU DisinfoLab, 28 September. www.disinfo.eu/publications/platforms-policies-on-ai-manipulated-and-generated-misinformation
- Milan, S., & Treré, E. (2019). Big data from the South(s): Beyond data universalism. *Television & New Media*, 20(4), 319–335. <https://doi.org/10.1177/1527476419837739>
- Milan, S., & van der Velden, L. (2016). The alternative epistemologies of data activism. *Digital Culture & Society*, 2(2), 57–74. <https://doi.org/10.14361/dcs-2016-0205>
- Milan, S., Treré, E., & Masiero, S. (Eds) (2021). *COVID-19 from the Margins: Pandemic Invisibilities, Policies and Resistance in the Datafied Society*. Institute of Network Cultures, Amsterdam University.
- Milmo, D. (2024). OpenAI putting ‘shiny products’ above safety, says departing researcher. *The Observer*, 28 May. www.theguardian.com/technology/article/2024/may/18/openai-putting-shiny-products-above-safety-says-departing-researcher
- Milton, v. c., & Mano, W. (2022). Afrokology and the right to communicate in Africa. *Javnost – The Public*, 29(1), 33–49. <https://doi.org/10.1080/13183222.2021.1889830>
- Misra, S. (2022). The overnment wants to sell your data | #SaveOurPrivacy. Internet Freedom Foundation, 5 March. <https://internetfreedom.in/the-government-wants-to-sell-your-data>
- MIT Technology Review Insights. (2023). Finding value in generative AI for financial services. MIT Technology Review and UBS. www.technologyreview.com/2023/11/26/1083841/finding-value-in-generative-ai-for-financial-services
- Mitchell, S., Potash, E., Barocas, S., D’Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, 8(2021), 141–163. <https://doi.org/10.1146/annurev-statistics-042720-125902>
- Mitchelstein, E., & Boczkowski, P. J. (Eds) (2023). *Digital Journalism in Latin America*. Routledge.

- Mitchelstein, E., Matassi, M., & Boczkowski, P. J. (2020). Minimal effects, maximum panic: Social media and democracy in Latin America. *Social Media + Society*, 6(4), 1–11. <https://doi.org/10.1177/2056305120984452>
- Mittelstadt, B., Wachter, S., & Russell, C. (2023). To protect science, we must use LLMs as zero-shot translators. *Nature Human Behaviour*, 7(11), 1830–1832. <https://doi.org/10.1038/s41562-023-01744-0>
- Mökander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2023). Auditing large language models: A three-layered approach. *AI and Ethics*, 2023, 1–31. <https://doi.org/10.1007/s43681-023-00289-2>
- Molina, M. D., & Sundar, S. S. (2022). Does distrust in humans predict greater trust in AI? Role of individual differences in user responses to content moderation. *New Media & Society*, 26(6), 3638–3656. <https://doi.org/10.1177/14614448221103534>
- Möller, J., Hameleers, M., & Ferreau, F. (2020). *Types of Disinformation and Misinformation: Various Types of Disinformation and Their Dissemination from a Communication Science and Legal Perspective*. Committee Chairperson Conference (GVK, Gremienvorsitzendenkonferenz) of the German Media Authorities. www.die-medienanstalten.de/fileadmin/user_upload/die_medienanstalten/Service/Studien_und_Gutachten/GVK_Summary_EN_final_web.pdf
- Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, 21(7), 959–977. <https://doi.org/10.1080/1369118X.2018.1444076>
- Monasterio Astobiza, A., Ausín, T., Liedo, B., Toboso, M., Aparicio, M., & López, D. (2022). Ethical Governance of AI in the Global South: A Human Rights Approach to Responsible Use of AI. *Proceedings* 81(136), 1–5. <https://doi.org/10.3390/proceedings2022081136>
- Mont’Alverne, C., Badrinathan, S., Ross Arguedas, A., Toff, B., Fletcher, R., & Nielsen, R. (2022). *The Trust Gap: How and Why News on Digital Platforms is Viewed More Sceptically Versus News in General*. Reuters Institute for the Study of Journalism, University of Oxford.
- Mont’Alverne, C., Ross Arguedas, A., Badrinathan, S., Toff, B., Fletcher, R., & Nielsen, R. K. (2023). Domain-specific influence on Facebook: How topic matters when assessing influential accounts in four countries. *Journal of Quantitative Description: Digital Media*, 3, 1–24. <https://ora.ox.ac.uk/objects/uuid:0e0ae80e-4bef-406f-a0b5-ef7e7cb6eaa9>
- Mont’Alverne, C., Ross Arguedas, A., Banerjee, S., Toff, B., Fletcher, R., & Nielsen, R. (2024). The electoral misinformation nexus: How news consumption, platform use, and trust in news influence belief in electoral misinformation. *Public Opinion Quarterly*, 88(S1), 681–707. <https://doi.org/10.1093/poq/nfae019>
- Montero, J., & Finger, M. (2021). *The Rise of the New Network Industries: Regulating Digital Platforms*. Routledge.
- Monti, M. (2020). The EU Code of Practice on Disinformation and the Risk of the Privatisation of Censorship. In S. Giusti & E. Piras (Eds), *Democracy and Fake News: Information Manipulation and Post-Truth Politics* (pp. 214–225). Routledge.
- Monzer, C., Moeller, J., Helberger, N., & Eskens, S. (2020). User perspectives on the news personalisation process: Agency, trust and utility as building blocks. *Digital Journalism*, 8(9), 1142–1162. <https://doi.org/10.1080/21670811.2020.1773291>
- Moore, M., & Tambini, D. (Eds) (2018). *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*. Oxford University Press.
- Moore, M., & Tambini, D. (Eds) (2021). *Regulating Big Tech: Policy Responses to Digital Dominance*. Oxford University Press.
- Moore, R. C., & Hancock, J. T. (2022). A digital media literacy intervention for older adults improves resilience to fake news. *Nature Scientific Reports*, 12(1), 1–9. <https://doi.org/10.1038/s41598-022-08437-0>
- Moran, C. R., Burton, J., & Christou. (2023). The US intelligence community, global security and AI: From secret intelligence to smart spying. *Journal of Global Security Studies*, 8(2), 1–18. <https://academic.oup.com/jogss/article/8/2/ogad005/7128314>
- Moran, R. E., & Shaikh, S. J. (2022). Robots in the news and newsrooms: Unpacking Meta-journalistic discourse on the use of artificial intelligence in journalism. *Digital Journalism*, 10(10), 1756–1774. <https://doi.org/10.1080/21670811.2022.2085129>
- Moreno Bellosio, N., & Petit, N. (2023). The EU Digital Markets Act (DMA): A competition hand in a regulatory glove. *European Law Review*, 48(4), 391–421. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4411743
- Morkūnas, M. (2023). Russian disinformation in the Baltics: Does it really work? *Public Integrity*, 25(6), 599–613. <https://doi.org/10.1080/10999922.2022.2092976>
- Morozov, E. (2013). *To Save Everything, Click Here: The Folly of Technological Solutionism*. Public Affairs.
- Motamedi, M. (2024). Iran unveils plan for tighter internet rules to promote local platforms. Al Jazeera, 24 February. www.aljazeera.com/news/2024/2/24/iran-unveils-plan-for-tighter-internet-rules-to-promote-local-platforms
- Motlagh, N. Y., Khajavi, M., Sharifi, A., & Ahmadi, M. (2023). The impact of artificial intelligence on the evolution of digital education: A comparative study of OpenAI text generation tools including ChatGPT, Bing Chat, Bard, and Ernie. *ArXiv*. <https://arxiv.org/abs/2309.02029>
- Motta, M., Hwang, J., & Stecula, D. (2024). What goes down must come up? Pandemic-related misinformation search behavior during an unplanned Facebook outage. *Health Communication*, 39(10), 2041–2052. <https://doi.org/10.1080/10410236.2023.2254583>
- Mudde, C. (2019). *The Far Right Today*. Polity Press.
- Mouffe, C. (1966). Democracy, Power, and the ‘Political’. In S. Benhabib (Ed.), *Democracy and Difference: Contesting the Boundaries of the Political* (pp. 245–256). Princeton University Press.

- Mueller, M. (2024). Rethinking AI Governance: The Political Economy of the Digital Ecosystem. In Proceedings of the TPRC2024 Conference, The Research Conference on Communications, Information and Internet Policy. <https://ssrn.com/abstract=4911002>
- Mueller, M. (2024). The Justice Department sues Georgia Tech: A teachable moment in cybersecurity management. Internet Governance Project, 27 August. www.internetgovernance.org/2024/08/27/the-justice-department-sues-georgia-tech-a-teachable-moment-in-cybersecurity-management
- Mühlhoff, R., & Ruschmeier, H. (2024). Regulating AI with purpose limitation for models. *Journal of AI Law and Regulation*, 1(1), 24–39. <https://doi.org/10.21552/aire/2024/1/5>
- Mukerjee, S. (2016). Net neutrality, Facebook, and India's battle to #SaveTheInternet. *Communication and the Public*, 1(3), 356–361. <https://doi.org/10.1177/2057047316665850>
- Mukhudwana, R. F. (2021). The rise of peripheral actors in media regulation in South Africa: An entry of social media mob(s). *African Journalism Studies*, 42(4), 153–178. <https://doi.org/10.1080/23743670.2022.2032783>
- Müller, M., & Kettemann, M. C. (2024). European Approaches to the Regulation of Digital Technologies. In H. Werthner, C. Ghezzi, J. Kramer, J. Nida-Rümelin, et al. (Eds), *Introduction to Digital Humanism: A Textbook* (pp. 623–637). Springer Nature Switzerland.
- Mulligan, C. E. A., & Godsiff, P. (2023). Datalism and data monopolies in the era of AI: A research agenda. *arXiv:2307.08049v1*. <https://doi.org/10.48550/arXiv.2307.08049>
- Munoriyarwa, A., & Mare, A. (2022). Mainstreaming Surveillance Through the Biometrification of Everyday Life. In A. Munoriyarwa & A. Mare (Eds), *Digital Surveillance in Southern Africa: Policies, Politics and Practices* (pp. 141–156). Springer International Publishing.
- Muñoz, K. (2023). *The Transformative Role of AI in Reshaping Electoral Politics*. DGAP Memo, No. 4. Deutsche Gesellschaft für Auswärtige Politik e.V.
- Murgia, M. (2019). Google accused of secretly feeding personal data to advertisers. *The Financial Times*. www.ft.com/content/e3e1697e-ce57-11e9-99a4-b5ded7a7fe3f
- Murphy, H., Pooler, M., & Harris, B. (2024). Brazil imposes fine on Musk's X after service returns. *The Financial Times*, 19 September. www.ft.com/content/2133cca8-ff48-4e11-8ca7-bd35d69201c5
- Murtaza, G., & Salman, S. (2019). GDPR fine imposed upon Google: An analysis. Manuscript. Brown University. <https://cs.brown.edu/courses/csci2390/2019/assign/gdpr/ssalman1-gmurtaza-google.pdf>
- Musiani, F. (2022). Infrastructuring digital sovereignty: A research agenda for an infrastructure-based sociology of digital self-determination practices. *Information, Communication & Society*, 25(6), 785–800. <https://doi.org/10.1080/1369118X.2022.2049850>
- Musoni, M., Karkare, P., & Teevan, C. (2024). *Cross-Border Data Flows in Africa: Continental Ambitions and Political Realities*. Discussion Paper No. 379. ecdpm (European Centre for Development Policy Management). <https://ecdpm.org/work/cross-border-data-flows-africa-continental-ambitions-and-political-realities>
- Mutsvairo, B. (2019). Challenges facing development of data journalism in non-Western societies. *Digital Journalism*, 7(9), 1289–1294. <https://doi.org/10.1080/21670811.2019.1691927>
- Mutsvairo, B., & Bebawi, S. (2019). Journalism educators, regulatory realities, and pedagogical predicaments of the 'fake news' era: A comparative perspective on the Middle East and Africa. *Journalism & Mass Communication Educator*, 74(2), 143–157. <https://doi.org/10.1177/1077695819833552>
- Mutsvairo, B., & Ragnedda, M. (Eds) (2019). *Mapping Digital Divide in Africa: A Mediated Analysis*. Amsterdam University Press.
- Mutsvairo, B., & Rønning, H. (2020). The Janus face of social media and democracy? Reflections on Africa. *Media, Culture & Society*, 42(3), 317–328. <https://doi.org/10.1177/0163443719899615>
- Myers, M., & Gilberds, H. (2024). *Are Donors Taking the Journalism Crisis Seriously? An Analysis of Official Aid to Media 2010–2019*. Center for International Media Assistance, National Endowment for Democracy, US.
- Naeem, S. B., Bhatti, R., & Khan, A. (2021). An exploration of how fake news is taking over social media and putting public health at risk. *Health Information & Libraries Journal*, 38(2), 143–149. <https://doi.org/10.1111/hir.12320>
- Nagaraj, A. (2020). Under watch: Indian city workers protest digital surveillance. Reuters, 17 March. www.reuters.com/article/idUSKBN21404S
- Nahmias, Y., & Perel, M. (2021). The oversight of content moderation by AI: Impact assessments and their limitations. *Harvard Journal on Legislation*, 58(1), 145–194. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3565025
- Nannini, L., Bonel, E., Bassi, D., & Maggini, M. J. (2024). Beyond phase-in: Assessing impacts on disinformation of the EU Digital Services Act. *AI and Ethics*, 1–29. <https://doi.org/10.1007/s43681-024-00467-w>
- Napoli, P. M. (2003). *Audience Economics: Media Institutions and the Audience Marketplace*. Columbia University Press.
- Napolitano, A. (2023). *Artificial Intelligence: The New Frontier of the EU's Border Externalisation Strategy*. EuroMed Rights.
- Nardi, B. A., & O'Day, V. (1999). *Information Ecologies: Using Technology with Heart*. MIT Press.
- Ncube, L. (2019). Digital media, fake news and pro-Movement for Democratic Change (MDC) alliance cyber-propaganda during the 2018 Zimbabwe Election. *African Journalism Studies*, 40(4), 44–61. <https://doi.org/10.1080/23743670.2019.1670225>

- Ndemo, B., & Thegeya, A. (2023). A Prototype Data Governance Framework for Africa. In B. Ndemo, N. Ndung'u, S. Odhiambo, & A. Shimeles (Eds), *Data Governance and Policy in Africa* (pp. 9–29). Springer.
- NDI (National Democratic Institute). (2023). *Mitigating Online Misleading Information and Polarization in Conflict-Sensitive Contexts: Experimental Evidence from Côte d'Ivoire*. www.ndi.org/publications/mitigating-online-misleading-information-and-polarization-conflict-sensitive-contexts
- NDIA (National Digital Inclusion Alliance). (2023). About NDIA. www.digitalinclusion.org/about-ndia
- Nechushtai, E. (2018). Could digital platforms capture the media through infrastructure? *Journalism*, 19(8), 1043–1058. <https://doi.org/10.1177/1464884917725163>
- Nechushtai, E. (2023). Resisting the individualization of risk: Strategies of engagement and caution in journalists' responses to online mobs in the United States and Germany. *Digital Journalism*, 11(10), 1906–1923. <https://doi.org/10.1080/21670811.2022.2153710>
- Neff, T., & Pickard, V. (2024). Funding democracy: Public media and democratic health in 33 countries. *The International Journal of Press/Politics*, 29(3), 601–627. <https://doi.org/10.1177/19401612211060255>
- Nenadić, I., Brogi, E., & Bleyer-Simon, K. (2023). *Structural Indicators to Assess Effectiveness of the EU's Code of Practice on Disinformation*. RSC Working Paper 2023/34. European University Institute.
- Nettle, D. (2023). The political economy of scientific publishing, and the promise of diamond open access. 14 June. www.danielnettle.org.uk/2023/06/14/the-political-economy-of-scientific-publishing-and-the-promise-of-diamond-open-access
- Neubaum, G., & Weeks, B. (2023). Computer-mediated political expression: A conceptual framework of technological affordances and individual tradeoffs. *Journal of Information Technology & Politics*, 20(1), 19–33. <https://doi.org/10.1080/19331681.2022.2028694>
- Newman, N., & Fletcher, R. (2017). *Bias, Bullshit and Lies: Audience Perspectives on Low Trust in the Media: Digital News Project 2017*. Reuters Institute for the Study of Journalism, University of Oxford.
- Newman, N., Fletcher, R., Robertson, C. T., Kirsten, E., & Nielsen, R. K. (2022). *Digital News Report 2022*. Reuters Institute for the Study of Journalism, University of Oxford. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022>
- Newman, N., Fletcher, R., Eddy, K., Robinson, C. T., & Nielsen, R. K. (2023). *Digital News Report 2023*. Reuters Institute for the Study of Journalism, University of Oxford. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023>
- Newman, N., Fletcher, R., Craig, T. R., Ross Arguedas, A., & Nielsen, R. K. (2024). *Digital News Report 2024*. Reuters Institute for the Study of Journalism, University of Oxford. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024>
- Newton, C., & Schiffer, Z. (2024). The Stanford Internet Observatory is being dismantled. Platformer, 13 June. www.platformer.news/stanford-internet-observatory-shutdown-stamos-diresta-sio
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology*, 58(1), 504–509. <https://doi.org/10.1002/pa2.487>
- Nguyễn, S., Kuo, R., Reddi, M., Li, L., & Moran, R. E. (2022). Studying mis- and disinformation in Asian diasporic communities: The need for critical transnational research beyond Anglocentrism. *Harvard Kennedy School (HKS) Misinformation Review*, 3(2), 1–12. <https://doi.org/10.37016/mr-2020-95>
- Nicas, J., & Conger, K. (2024). Brazil blocks X after Musk ignores court orders. *The New York Times*, 30 August. www.nytimes.com/2024/08/30/world/americas/brazil-elon-musk-x-blocked.html
- Nicholas, G., & Bhatia, A. (2023). *Lost in Translation: Large Language Models in Non-English Content Analysis*. Center for Democracy & Technology, US & Belgium.
- Nicholson, C. (2024). Elsevier parent reports 10% hike in profits for 2023. *Research Professional News*, 15 February. www.researchprofessionalnews.com/rr-news-europe-infrastructure-2024-2-elsevier-parent-reports-10-hike-in-profits-for-2023
- Nicoli, N., & Iosifidis, P. (2023). EU digital economy competition policy: From ex-post to ex-ante. The case of Alphabet, Amazon, Apple, and Meta. *Global Media and China*, 8(1), 24–38. <https://doi.org/10.1177/20594364231152673>
- Nieborg, D. B., & Poell, T. (2018). The platformization of cultural production: Theorizing the contingent cultural commodity. *New Media & Society*, 20(11), 4275–4292. <https://doi.org/10.1177/1461444818769694>
- Nieborg, D., Poell, T., Caplan, R., & van Dijck, J. (2024). Locating and theorising platform power. *Internet Policy Review*, 13(2), 1–17. <https://policyreview.info/articles/analysis/introduction-special-issue-locating-and-theorising-platform-power>
- Nielsen, R. K., & Fletcher, R. (2023). Comparing the platformization of news media systems: A cross-country analysis. *European Journal of Communication*, 38(5), 484–499. <https://doi.org/10.1177/02673231231189043>
- Nielsen, R. K., & Ganter, S. A. (2022). *The Power of Platforms: Shaping Media and Society*. Oxford University Press.
- Nielsen, R. K., Gorwa, R., & de Cock Buning, M. (2019). *What Can Be Done? Digital Media Policy Options for Strengthening European Democracy*. Reuters Institute for the Study of Journalism, University of Oxford.
- Nieminen, H. (2024). Why We Need Epistemic Rights. In M. A. Horowitz, H. Nieminen, K. Lehtisaari, & A. D'Arma (Eds), *Epistemic Rights in the Era of Digital Disruption* (pp. 11–28). Springer International Publishing.
- Niklas, J., & Dencik, L. (2024). Data justice in the 'twin objective' of market and risk: How discrimination is formulated in EU's AI policy. *Policy & Internet*, 16(3), 509–522. <https://doi.org/10.1002/poi3.392>

- Ning, L., Liu, L., Wu, J., Wu, N., Berlowitz, D., Prakash, S., Green, B., O'Banion, S., & Xie, J. (2024). *User-LLM: Efficient LLM contextualization with user embeddings*. arXiv. <https://doi.org/10.48550/arXiv.2402.13598>
- Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus*, 140(4), 32–48. <https://direct.mit.edu/daed/article/140/4/32/26914/A-Contextual-Approach-to-Privacy-Online>
- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.
- Nordic Council of Ministers Secretariat. (2023). *A Nordic Approach to Democratic Debate in the Age of Big Tech : Recommendations from the Nordic Think Tank for Tech and Democracy*. <http://dx.doi.org/10.6027/nord2023-004>
- Nothias, T. (2020). Access granted: Facebook's free basics in Africa. *Media, Culture & Society*, 42(3), 329–348. <https://doi.org/10.1177/0163443719890530>
- Notley, T., Dezuanni, M., Zhong, H., & Chambers, S. (2020). *News and Young Australians in 2020: How Young People Access, Perceive and Are Affected by News Media*. Western Sydney University and QUT Digital Media Research Centre.
- Novet, J. (2023). Microsoft's \$13 billion bet on OpenAI carries huge potential along with plenty of uncertainty. CNBC, 8 April. www.cnbc.com/2023/04/08/microsofts-complex-bet-on-openai-brings-potential-and-uncertainty.html
- Nowotny, H. (2021). *In AI We Trust: Power, Illusion and Control of Predictive Algorithms*. Polity Press.
- NSF (National Science Foundation). (2023). *NSF Public Access Plan 2.0: Ensuring Open, Immediate and Equitable Access to National Science Foundation Funded Research*. www.nsf.gov/pubs/2023/nsf23104/nsf23104.pdf
- Nyalety, E., Parizi, R. M., Zhang, Q., & Choo, K. K. R. (2019). BlockIPFS – Blockchain-enabled InterPlanetary File System for Forensic and Trusted Data Traceability. 2019 IEEE International Conference on Blockchain, 1 July. doi: 10.1109/Blockchain.2019.00012.
- Nyamnjoh, F. B. (2006). *Insiders and Outsiders: Citizenship and Xenophobia in Contemporary Southern Africa*. Zed Books.
- Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., et al. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620, 137–144. <https://doi.org/https://doi.org/10.1038/s41586-023-06297-w>
- NYPD (New York City Police Department). (2024). Public Oversight of Surveillance Technology (POST) Act impact and use policies. www.nyc.gov/site/nypd/about/about-nypd/policy/post-act.page
- Ó Fathaigh, R., Helberger, N., & Appelman, N. (2021). The perils of legally defining disinformation. *Internet Policy Review*, 10(4), 1–25. <https://policyreview.info/articles/analysis/perils-legally-defining-disinformation>
- Ó Siochrú, S., Singh, P. J., & Burch, S. (2024). Just Net Coalition: Seeking Anchors of Equity and Social Justice in a Digital World. In C. Padovani, V. Wavre, A. Hintz, G. Goggin, & P. Iosifidis (Eds), *Global Communication Governance at the Crossroads* (pp. 239–257). Palgrave Macmillan.
- O'Brien, I. (2024). Data center emissions probably 662% higher than big tech claims. Can it keep up the ruse? *The Guardian*, 15 September. www.theguardian.com/technology/2024/sep/15/data-center-gas-emissions-tech
- O'Brien, M., & Swenson, A. (2024). Tech companies sign accord to combat AI-generated election trickery. The Associated Press, 16 February. <https://apnews.com/article/ai-generated-election-deepfakes-munich-accord-meta-google-microsoft-tiktok-x-c40924ffc68c94fac74fa994c520fc06>
- O'Brien, M., & Swenson, A. (2024). Tech companies sign accord to combat AI-generated election trickery. The Associated Press, 16 February. <https://apnews.com/article/ai-generated-election-deepfakes-munich-accord-meta-google-microsoft-tiktok-x-c40924ffc68c94fac74fa994c520fc06>
- O'Connor, C. (2022). #Propaganda: Russia state-controlled media flood TikTok with Ukraine disinformation. ISD Digital Dispatches, 2 March. www.isdglobal.org/digital_dispatches/propaganda-russia-state-controlled-media-flood-tiktok-with-ukraine-disinformation
- O'Faolain, A. (2024). Court dismisses claim data watchdog failed to investigate alleged breach by Google. *The Irish Times*, 24 June. www.irishtimes.com/crime-law/courts/2024/06/24/court-dismisses-claim-data-watchdog-failed-to-investigate-alleged-breach-by-google
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Harvard University Press.
- O'Neill, O. (2002). *A Question of Trust: The BBC Reith Lectures 2002*. Cambridge University Press.
- Obreja, D. M. (2023). Mapping the political landscape on social media using bibliometrics: A longitudinal co-word analysis on Twitter and Facebook publications published between 2012 and 2021. *Social Science Computer Review*, 41(5), 1712–1728. <https://doi.org/10.1177/0894439322111749>
- OECD (Organisation for Economic Co-operation and Development) (2022a). *Going Digital to Advance Data Governance for Growth and Well-Being*. <https://doi.org/10.1787/e3d783b0-en>
- OECD (Organisation for Economic Co-operation and Development). (2022b). *Misinformation and Disinformation: An International Effort Using Behavioural Science to Tackle the Spread of Misinformation*. OECD Public Governance Policy Papers No. 21. <https://doi.org/10.1787/b7709d4f-en>
- OECD (Organisation for Economic Co-operation and Development). (2022c). *Recommendation of the Council on Artificial Intelligence*. <https://oecd.ai/en/assets/files/OECD-LEGAL-0449-en.pdf>
- OECD. (2023). Catalogue of tools & metrics for trustworthy AI. <https://oecd.ai/en/catalogue>
- OECD. (2024). *Facts not Fakes: Tackling Disinformation, Strengthening Information Integrity*. www.oecd.org/en/publications/facts-not-fakes-tackling-disinformation-strengthening-information-integrity_d909ff7a-en.html

- Ofcom. (2022). *Adults' Media Use and Attitudes Report 2022*. www.ofcom.org.uk/media-use-and-attitudes/media-habits-children/children-and-parents-media-use-and-attitudes-report-2022
- Ofcom. (2023a). *Adults' Media Use and Attitudes Report 2023*. www.ofcom.org.uk/siteassets/resources/documents/research-and-data/media-literacy-research/adults/adults-media-use-and-attitudes-2023/adults-media-use-and-attitudes-report-2023.pdf?v=329409
- Ofcom. (2023b). *Media Literacy by Design: Best Practice Principles for On-Platform Interventions to Promote Media Literacy*. www.ofcom.org.uk/siteassets/resources/documents/consultations/category-2-6-weeks/270395---media-literacy-by-design/secondary-documents/call-for-input-best-practice-principles-for-media-literacy?v=330332
- Ofcom. (2023c). *News Consumption in the UK: 2023 Research Findings*. https://www.ofcom.org.uk/_data/assets/pdf_file/0024/264651/news-consumption-2023.pdf
- Ofcom. (2023d). *Online Nation 2023 Report*. www.ofcom.org.uk/media-use-and-attitudes/online-habits/online-nation
- Ofcom. (2024a). *Children's Media Lives 2024: Ten Years of Longitudinal Research, A Report by Ofcom*. www.ofcom.org.uk/media-use-and-attitudes/media-habits-children/childrens-media-lives
- Ofcom. (2024b). *Future Technology and Media Literacy: Understanding Generative AI*. www.ofcom.org.uk/media-use-and-attitudes/media-literacy/discussion-papers/?language=en
- Offenhuber, D. (2024). Shapes and frictions of synthetic data. *Big Data & Society*, 11(2), 1–16. <https://doi.org/10.1177/20539517241249390>
- OHCHR (Office of the United Nations High Commissioner for Human Rights). (2021). *Disinformation and Freedom of Opinion and Expression*. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan, A/HRC/47/25. www.ohchr.org/en/documents/thematic-reports/ahrc4725-disinformation-and-freedom-opinion-and-expression-report
- OHCHR (Office of the United Nations High Commissioner for Human Rights). (1993). *Vienna Declaration and Programme of Action*. www.ohchr.org/en/instruments-mechanisms/instruments/vienna-declaration-and-programme-action
- OHCHR (Office of the United Nations High Commissioner for Human Rights). (2014). *Factors that Impede Equal Political Participation and Steps to Overcome those Challenges*. www.ohchr.org/en/documents/thematic-reports/ahrc2729-factors-impede-equal-political-participation-and-steps-overcome
- Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., et al. (2024). Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. *Communication Methods and Measures*, 18(2), 124–141. www.tandfonline.com/doi/full/10.1080/19312458.2023.2181319
- Okolo, C. T. (2023). AI in the Global South: Opportunities and challenges towards more inclusive governance. Brookings Institution Commentary, 1 November. www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challenges-towards-more-inclusive-governance
- Okunlaya, R. O., Syed Abdullah, N., & Alias, R. A. (2022). Artificial intelligence (AI) library services innovative conceptual framework for the digital transformation of university education. *Library Hi Tech*, 40(6), 1869–1892. <https://doi.org/10.1108/LHT-07-2021-0242>
- Okyere-Manu, B. (2023). The Place of the African Relational and Moral Theory of Ubuntu in the Global Artificial Intelligence and Big Data Discussion: Critical Reflections. In M. Findlay, L. M. Ong, & W. Zhang (Eds), *Elgar Companion to Regulating AI and Big Data in Emerging Economies* (pp. 99–114). Edward Elgar Publishing.
- Oladapo, O., & Ojebode, A. (2021). *Nigeria Digital Rights Landscape Report*. Institute for Development Studies, Sussex.
- Oliver, P., Lim, C., Matthews, M., & Hanna, A. (2022). Black protests in the United States, 1994 to 2010. *Sociological Science*, 9, 275–312. <https://doi.org/10.15195/v9.a12>
- Olorunju, N., & Adams, R. (2024). African data trusts: New tools towards collective data governance? *Information & Communications Technology Law*, 33(1), 85–98. <https://doi.org/10.1080/13600834.2023.2260678>
- Omar, N. (2023). Data, AI and elections in Africa: Working towards a collaborative African response. Research ICT Africa, South Africa, 22 November. <https://researchictafrica.net/2023/11/22/data-ai-and-elections-in-africa-working-towards-a-collaborative-african-response>
- Ong, J. C. L., Chang, S. Y.-H., William, W., Butte, A. J., et al. (2024). Ethical and regulatory challenges of large language models in medicine. *The Lancet Digital Health*, 6(6), e428–e432. [https://doi.org/10.1016/S2589-7500\(24\)00061-X](https://doi.org/10.1016/S2589-7500(24)00061-X)
- Ong, J. C., & Das, R. (2020). Two Concepts from Television Audience Research in Times of Datafication and Disinformation. In S. Shimpach (Ed.), *The Routledge Companion to Global Television* (Chapter 9). Routledge.
- Ong, J. C., & Tapsell, R. (2022). Demystifying disinformation shadow economies: Fake news work models in Indonesia and the Philippines. *Asian Journal of Communication*, 32(3), 251–267. <https://doi.org/10.1080/01292986.2021.1971270>
- OSCE (Organization for Security and Co-operation in Europe). (2017). *Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda*. OSCE, 3 March. www.osce.org/fom/302796
- Ozanne, M., Bhandari, A., Bazarova, N. N., & DiFranzo, D. (2022). Shall AI moderators be made visible? Perception of accountability and trust in moderation systems on social media platforms. *Big Data & Society*, 9(2), 1–13. <https://doi.org/10.1177/2053951722115666>
- Paal, B. (2017). Current issues and recent developments on media concentration in the context of competition law and media law. *Journal of Intellectual Property Law and Practice*, 12(7), 610–616. <https://academic.oup.com/jiplp/article-abstract/12/7/610/3916940>

- Padovani, C., Wavre, V., Hintz, A., Goggin, G., & Iosifidis, P. (Eds) (2024). *Global Communication Governance at the Crossroads*. Palgrave Macmillan.
- Page, P., Knowles, P., & Mitwicki, R. (2023). Distributed governance: A principal-agent approach to data governance – Part 1. Background & core definitions. *ArXiv*. <http://arxiv.org/abs/2308.07280>
- Palmer, R., & Toff, B. (2022). Neither absent nor ambient: Incidental news exposure from the perspective of news avoiders in the UK, United States, and Spain. *The International Journal of Press/Politics*, 29(3), 755–773. <https://doi.org/10.1177/19401612221103144>
- Palmer, R., Toff, B., & Nielsen, R. K. (2023). Examining assumptions around how news avoidance gets defined: The importance of overall news consumption, intention, and structural inequalities. *Journalism Studies*, 24(6), 697–714. <https://doi.org/10.1080/1461670X.2023.2183058>
- Pamment, J. (2020). *EU Code of Practice on Disinformation: Briefing Note for the New European Commission*. Working Paper. Carnegie Endowment for International Peace.
- Pan, C. A., Yakhmi, S., Iyer, T. P., Strasnick, E., Zhang, A. X., & Bernstein, M. S. (2022). Comparing the perceived legitimacy of content moderation processes: Contractors, algorithms, expert panels, and digital juries. *Proceedings of the ACM on Human-Computer Interaction*, 6(82), 1–31. <https://doi.org/10.1145/3512929>
- Papacharissi, Z. A. (2013). *A Private Sphere: Democracy in a Digital Age*. John Wiley & Sons.
- Papaevangelou, C. (2023). Funding Intermediaries: Google and Facebook's strategy to capture journalism. *Digital Journalism*, 12(2), 234–255. <https://doi.org/10.1080/21670811.2022.2155206>
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin.
- Park, G. (2020). The changing wind of data privacy law: A comparative study of the European Union's General Data Protection Regulation and the 2018 California Consumer Privacy Act. *UC Irvine Law Review*, 10(4), 1455–1489. <https://escholarship.org/content/qt8562f0v0/qt8562f0v0.pdf>
- Park, S., & Humphry, J. (2019). Exclusion by design: Intersections of social, digital and data exclusion. *Information, Communication & Society*, 22(9), 934–944. <https://doi.org/10.1080/1369118X.2019.1606266>
- Park, S., Strover, S., Choi, J., & Schnell, M. (2023). Mind games: A temporal sentiment analysis of the political messages of the Internet Research Agency on Facebook and Twitter. *New Media & Society*, 25(3), 463–484. <https://doi.org/10.1177/14614448211014355>
- Park, Y. J. (2024). Algorithmic bias, marketplaces, and diversity regulation. In Proceedings of the TPRC 2024 Conference, Washington DC. <https://ssrn.com/abstract=4912448>
- Pasek, J., Feldman, L., Romer, D., & Jamieson, K. H. (2008). Schools as incubators of democratic participation: Building long-term political efficacy with civic education. *Applied Developmental Science*, 12(1), 26–37. <https://doi.org/10.1080/10888690801910526>
- Pasquinelli, M. (2023). *The Eye of the Master: A Social History of Artificial Intelligence*. Verso Books.
- Patrick, S., & Huggins, A. (2023). The term 'Global South' is surging. It should be retired. *Carnegie Endowment for International Peace, Commentary*, 15 August. <https://carnegieendowment.org/posts/2023/08/the-term-global-south-is-surging-it-should-be-retired?lang=en>
- Paul, R., Carmel, E., & Cobbe, J. (Eds) (2024). *Handbook on Public Policy and Artificial Intelligence*. Edward Elgar Publishing.
- Payne, R., Martin, B. A. S., Tuzovic, S., & Wang, S. (2023). Defining biometrics with privacy and benefits: A research agenda. *Australasian Marketing Journal*, 31(4), 294–302. <https://doi.org/10.1177/14413582231167645>
- Pennycook, G., & Rand, D. G. (2022). Nudging social media toward accuracy. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 152–164. <https://doi.org/10.1177/00027162221092342>
- Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, 31(7), 770–780. <https://doi.org/10.1177/0956797620939054>
- Pentney, K. (2024). States' positive obligation to create a favourable environment for participation in public debate: A principle in search of a practical effect? *Journal of Media Law*, 16(1), 146–177. <https://doi.org/10.1080/17577632.2024.2326278>
- Pequeño IV, A. (2023). Twitter starts paying creators tens of thousands – Amid intensifying competition with Threads. *Forbes*, 13 July. www.forbes.com/sites/antoniopequenoiv/2023/07/13/twitter-starts-paying-creators-tens-of-thousands---amid-intensifying-competition-with-threads
- Pérez-Curiel, C., Rivas-de-Roca, R., & García-Gordillo, M. (2021). Impact of Trump's digital rhetoric on the US elections: A view from worldwide far-right populism. *Social Sciences*, 10(5), 1–17. <https://doi.org/10.3390/socsci10050152>
- Perrigo, B. (2023). Exclusive: OpenAI lobbied EU to water down AI regulation. *TIME*, 20 June. <https://time.com/6288245/openai-eu-lobbying-ai-act>
- Peters, J. (2016). Why is it so expensive to read academic research? *Slate*, 5 April. <https://slate.com/technology/2016/04/the-lawsuit-against-sci-hub-begs-the-question-why-are-academic-journals-so-expensive-anyway.html>
- Peters, J. (2023). X stops showing headlines because Elon Musk thinks it will make posts look better. *The Verge*, 4 October. www.theverge.com/2023/10/4/23903859/x-elon-musk-headlines-links-image-twitter

- Petre, C. (2021). *All the News That's Fit to Click: How Metrics Are Transforming the Work of Journalists*. Princeton University Press.
- Pfeiffer, J., Gutschow, J., Haas, C., Möslin, F., et al. (2023). Algorithmic fairness in AI. *Business & Information Systems Engineering*, 65(2), 209–222. <https://doi.org/10.1007/s12599-023-00787-x>
- Phan, T., Goldenfein, J., Mann, M., & Kuch, D. (2022). Economies of virtue: The circulation of 'ethics' in big tech. *Science as Culture*, 31(1), 121–135. <https://doi.org/10.1080/09505431.2021.1990875>
- Pickard, V. (2014). *America's Battle for Media Democracy: The Triumph of Corporate Libertarianism and the Future of Media Reform*. Cambridge University Press.
- Pickard, V. (2020a). *Democracy Without Journalism: Confronting the Misinformation Society*. Oxford University Press.
- Pickard, V. (2020b). Monopoly Control over Digital Infrastructures. In V. Pickard, *Democracy without Journalism? Confronting the Misinformation Society* (pp. 104–135). Oxford University Press.
- Pickard, V. (2020c). Restructuring democratic infrastructures: A policy approach to the journalism crisis. *Digital Journalism*, 8(6), 704–719. www.tandfonline.com/doi/full/10.1080/21670811.2020.1733433
- Pickard, V. (2022a). Can Journalism Survive in the Age of Platform Monopolies? Confronting Facebook's Negative Externalities. In T. Flew & F. R. Martin (Eds), *Digital Platform Regulation: Global Perspectives on Internet Governance* (pp. 23–42). Palgrave Macmillan.
- Pickard, V. (2022b). Democratizing the platforms: Promises and perils of public utility regulation. *Media Development*, 68(3), 6–10. www.asc.upenn.edu/sites/default/files/2022-10/Democratizing%20the%20platforms%20-%20Victor%20Pickard.pdf
- Pickard, V. (2024). Legal foundations for non-reformist media reforms. Knight First Amendment Institute, Columbia University, 16 July. <http://knightcolumbia.org/content/legal-foundations-for-non-reformist-media-reforms>
- Pickard, V., & Berman, D. E. (2019). *After Net Neutrality: A New Deal for the Digital Age*. Yale University Press.
- Pielemeier, J. (2020). Disentangling disinformation: What makes regulating disinformation so difficult? Symposium: News, disinformation, and social media responsibility. *Utah Law Review*, 2020(4), 917–940. <https://doi.org/10.26054/OD-CJBV-FTGJ>
- Pinkleton, B. E., Austin, E. W., Zhou, Y., Willoughby, J. F., & Reiser, M. (2012). Perceptions of news media, external efficacy, and public affairs apathy in political decision making and disaffection. *Journalism & Mass Communication Quarterly*, 89(1), 23–39. <https://doi.org/10.1177/1077699011428586>
- Pintak, L., & Ginges, J. (2008). The mission of Arab journalism: Creating change in a time of turmoil. *The International Journal of Press/Politics*, 13(3), 193–227. <https://doi.org/10.1177/1940161208317142>
- Pintak, L., & Nazir, S. J. (2013). Pakistani journalism: At the crossroads of Muslim identity, national priorities and journalistic culture. *Media, Culture & Society*, 35(5), 640–665. <https://doi.org/10.1177/0163443713483654>
- Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, 20(1), 293–310. <https://doi.org/10.1177/1461444816661553>
- Pocino, P. V. (2021). *Algorithms in the Newsrooms: Challenges and Recommendations for Artificial Intelligence with the Ethical Values of Journalism*. Catalan Press Council.
- Poell, T., Nieborg, D. B., & Duffy, B. E. (2023). Spaces of negotiation: Analyzing platform power in the news industry. *Digital Journalism*, 11(8), 1391–1409. <https://doi.org/10.1080/21670811.2022.2103011>
- Poell, T., Nieborg, D., & van Dijck, J. (2019). Platformisation. *Internet Policy Review*, 8(4), 1–13. <https://policyreview.info/concepts/platformisation>
- Pollicino, O., & De Gregorio, G. (2022). Constitutional democracy, platform powers and digital populism. *Constitutional Studies*, 8(1), 11–34. <https://constitutionalstudies.wisc.edu/index.php/cs/article/view/87>
- Polyák, G. (2019). Media in Hungary: Three Pillars of an Illiberal Democracy. In E. Połńska & C. Beckett (Eds), *Public Service Broadcasting and Media Systems in Troubled European Democracies* (pp. 279–303). Palgrave Macmillan.
- Popiel, P. (2018). The Tech Lobby: Tracing the contours of new media elite lobbying power. *Communication, Culture and Critique*, 11(4), 566–585. <https://doi.org/10.1093/ccc/tcy027>
- Porter, E., & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *PNAS: Proceedings of the National Academy of Sciences*, 118(37), 1–7. <https://doi.org/10.1073/pnas.2104235118>
- Posada, A., & Chen, G. (2018). Inequality in knowledge production: The integration of academic infrastructure by big publishers. *22nd International Conference on Electronic Publishing*. <https://pdfs.semanticscholar.org/4bb0/2d560d5d7b74560b9f9a6a6fe1493150f393.pdf>
- Posetti, J., & Bontcheva, K. (2020). *Disinfodemic: Deciphering COVID-19 Disinformation*. UNESCO Policy Brief 1. <https://unesdoc.unesco.org/ark:/48223/pf0000374416>
- Posetti, J., Shabbir, N., Douglas, O., & Gardiner, B. (2022). *The Chilling: A Global Study of Online Violence Against Women Journalists*. ICFJ (International Center for Journalists), UNESCO.
- Pospieszna, P., Lown, P., & Dietrich, S. (2023). Building active youth in post-Soviet countries through civic education programmes: Evidence from Poland. *East European Politics*, 39(2), 321–344. <https://doi.org/10.1080/21599165.2022.2102608>

- Post, R. (2009). Hate Speech. In I. Hare & J. Weinstein (Eds), *Extreme Speech and Democracy* (pp. 123–138). Oxford University Press.
- Powell, A. B. (2015). Open culture and innovation: Integrating knowledge across boundaries. *Media, Culture & Society*, 37(3), 376–393. <https://doi.org/10.1177/0163443714567169>
- Powell, A. B. (2021). *Undoing Optimization: Civic Action in Smart Cities*. Yale University Press.
- Pradhan, S. (2021). Combating online misinformation: Diverging approaches in the United States and European Union. *Columbia Journal of Transnational Law Bulletin*, 15 April. www.jtl.columbia.edu/bulletin-blog/combating-online-misinfo
- Prasad, R. (2018). Ascendant India, digital India: How net neutrality advocates defeated Facebook’s Free Basics. *Media, Culture & Society*, 40(3), 415–431. <https://doi.org/10.1177/0163443717736117>
- Prior, M. (2005). News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout. *American Journal of Political Science*, 49(3), 577–592. <https://doi.org/10.1111/j.1540-5907.2005.00143.x>
- Prochazka, F., & Schweiger, W. (2019). How to measure generalized trust in news media? An adaptation and test of scales. *Communication Methods and Measures*, 13(1), 26–42. <https://doi.org/10.1080/19312458.2018.1506021>
- Procter, R., Arana-Catania, M., He, Y., Liakata, M., et al. (2023). Some observations on fact-checking work with implications for computational support. *ArXiv*. <https://arxiv.org/pdf/2305.02224>
- Procter, R., Webb, H., Jirotko, M., Burnap, P., et al. (2019). A study of cyber hate on Twitter with implications for social media governance strategies. *ArXiv*. <https://doi.org/10.48550/arXiv.1908.11732>
- Program for Public Consultation. (2022). Three-in-four voters favor reinstating net neutrality. School of Public Policy, University of Maryland. <https://publicconsultation.org/united-states/three-in-four-voters-favor-reinstating-net-neutrality>
- PublicSpaces International. (2024). How we think about AI is largely dictated by Big Tech. <https://english.publicspaces.net/2024/06/03/how-we-think-about-ai-is-largely-dictated-by-big-tech>
- Puddephatt, A. (2021). *Letting the Sun Shine In: Transparency and Accountability in the Digital Age*. UNESCO.
- Puehringer, S., Rath, J., & Griesebner, T. (2021). *The political economy of academic publishing*. *PLOS ONE*, 16(6), 1–21. <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0253226&type=printable>
- Punia, S., Mohan, S., Kakkar, J. M., & Bhandari, V. (Eds) (2022). *Emerging Trends in Data Governance*. National Law University, Delhi Press.
- Puppis, M., Mansell, R., & Van den Bulck, H. (Eds) (2024). *Handbook of Media and Communication Governance*. Edward Elgar Publishing.
- Purandare, U., & Parkar, K. (2020). ‘Eyes and Ears’: Surveillance in the Indian Smart City. In J. C. Augusto (Ed.), *Handbook of Smart Cities* (pp. 1–32). Springer International Publishing.
- Purtova, N., & van Maanen, G. (2024). Data as an economic good, data as a commons, and data governance. *Law, Innovation and Technology*, 16(1), 1–42. <https://doi.org/10.1080/17579961.2023.2265270>
- Pyo, J. Y. (2022). Different stakes, different struggles, and different practices to survive: News organizations and the spectrum of platform dependency. *New Media & Society*, 26(8), 4572–4588. <https://doi.org/10.1177/14614448221123279>
- Quelle, D., & Bovet, A. (2024). The perils and promises of fact-checking with large language models. *Frontiers in Artificial Intelligence*, 7(1341697), 1–14. <https://doi.org/10.3389/frai.2024.1341697>
- Quintais, J. P., De Gregorio, G., & Magalhães, J. C. (2023). How platforms govern users’ copyright-protected content: Exploring the power of private ordering and its implications. *Computer Law & Security Review*, 48(2023), 1–25. <https://doi.org/10.1016/j.clsr.2023.105792>
- Quintavalla, A., & Temperman, J. (Eds) (2023). *Artificial Intelligence and Human Rights*. Oxford University Press.
- Radsch, C. C. (2022). *Making Big Tech Pay for the News They Use*. Center for International Media Assistance, US.
- Radsch, C. C. (2023a). Envisioning a healthy information ecosystem. CDT (Center for Democracy and Technology), 2 June. <https://cdt.org/insights/from-our-fellows-envisioning-a-healthy-information-ecosystem>
- Radsch, C. C. (2023b). Platformization and media capture: A framework for regulatory analysis of media-related platform regulations. *UCLA Journal of Law & Technology*, 28(2), 175–223. <https://uclajolt.com/platformization-and-media-capture-a-framework-for-regulatory-analysis-of-media-related-platform-regulations>
- Radsch, C. C. (2023c). The value of news content to Google is way more than you think. *Tech Policy Press*, 8 August. <https://techpolicy.press/the-value-of-news-content-to-google-is-way-more-than-you-think>
- Radsch, C. C. (2023d). *URGENT: Understanding and Responding to Global Emerging News Threats*. Internews. <https://internews.org/wp-content/uploads/2023/03/29Mar-URGENT-Report-Final.pdf>
- Radsch, C. C. (2023e). What makes for a healthy information ecosystem? New visual tool. <https://taicollaborative.org/what-makes-for-a-healthy-information-ecosystem-new-visual-tool>
- Radsch, C. C. (2024). The must-carry solution for the media’s Google problem. *Washington Monthly*, 22 August. <https://washingtonmonthly.com/2024/08/22/the-must-carry-solution-for-the-medias-google-problem>
- Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *ACM Conference on Fairness, Accountability, and Transparency (FAT*) 2020*, 469–481. <http://dx.doi.org/10.2139/ssrn.3408010>
- Ragnedda, M., & Gladkova, A. (Eds) (2020). *Digital Inequalities in the Global South*. Springer Nature.

- Ragnedda, M., & Ruiu, M. L. (2020). *Digital Capital: A Bourdieusian Perspective on the Digital Divide*. Emerald Group Publishing.
- Rajkumar, M., & Ashraf, M. M. (2023). *Response to Draft Amendment to the IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021*. IT for Change, India.
- Ramiro, A., & Cruz, L. (2023). The grey-zones of public-private surveillance: Policy tendencies of facial recognition for public security in Brazilian cities. *Internet Policy Review*, 12(1), 1–28. <https://policyreview.info/articles/analysis/grey-zones-public-private-surveillance-policy-tendencies-facial-recognition>
- Ranaivoson, H., Afilipoaie, A., & Domazetovikj, N. (2022). *Media Pluralism in the EU: A Prospective Look at the European Media Freedom Act*. Policy Brief No. 64. Vrije Universiteit Brussel, Studies in Media Innovation Technology Research Group.
- Rankin, J. (2023). MEPs launch site for EU officials to report 'shady lobbying' by big tech. *The Guardian*, 2 February. <https://amp.theguardian.com/world/2023/feb/02/meps-launch-site-for-eu-officials-to-report-shady-lobbying>
- Rantanen, T. (2024). *Dead Men's Propaganda: Ideology and Utopia in Comparative Communications Studies*. LSE Press.
- Ratner, Y., Dvir Gvirsman, S., & Ben-David, A. (2023). 'Saving journalism from Facebook's death grip'? The implications of content-recommendation platforms on publishers and their audiences. *Digital Journalism*, 11(8), 1410–1431. <https://doi.org/10.1080/21670811.2023.2180403>
- Recabarren, R., Carbutar, B., Hernandez, N., & Shafin, A. A. (2023, 2023). *Strategies and Vulnerabilities of Participants in Venezuelan Influence Operations*. 32nd USENIX Security Symposium. www.usenix.org/conference/usenixsecurity23/presentation/recabarren
- Recuero, R. (2024). The platformization of violence: Toward a concept of discursive toxicity on social media. *Social Media + Society*, 10(1), 1–9. <https://doi.org/10.1177/20563051231224264>
- Recuero, R., Soares, F. B., & Gruz, A. (2020). Hyperpartisanship, Disinformation and Political Conversations on Twitter: The Brazilian Presidential Election of 2018. In AAAI (Association for the Advancement of Artificial Intelligence) (Ed.), *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 569–578).
- Recuero, R., Soares, F. B., Vinhas, O., Volcan, T., Hüttner, L. R. G., & Silva, V. (2022). Bolsonaro and the far right: How disinformation about COVID-19 circulates on Facebook in Brazil. *International Journal of Communication*, 16(2022), 148–171. <https://ijoc.org/index.php/ijoc/article/viewFile/17724/3631>
- Reddi, M., Kuo, R., & Kreiss, D. (2023). Identity propaganda: Racial narratives and disinformation. *New Media & Society*, 25(8), 2201–2218. <https://doi.org/10.1177/14614448211029293>
- Reinemann, C., Stanyer, J., Scherr, S., & Legnante, G. (2012). Hard and soft news: A review of concepts, operationalizations and key findings. *Journalism*, 13(2), 221–239. <https://doi.org/10.1177/1464884911427803>
- Reis, O., Eneh, N. E., Ehimuan, B., Anyanwu, A., Olorunsogo, T., & Abrahams, T. O. (2024). Privacy law challenges in the digital age: A global review of legislation and enforcement. *International Journal of Applied Research in Social Sciences*, 6(1), 73–88. <https://doi.org/10.51594/ijarss.v6i1.733>
- Reisach, U. (2021). The responsibility of social media in times of societal and political manipulation. *European Journal of Operational Research*, 291(3), 906–917. <https://doi.org/10.1016/j.ejor.2020.09.020>
- Reiter, F., & Matthes, J. (2023). Correctives of the mainstream media? A panel study on mainstream media use, alternative digital media use, and the erosion of political interest as well as political knowledge. *Digital Journalism*, 11(5), 813–832. <https://doi.org/10.1080/21670811.2021.1974916>
- Renaldi, A. (2024). In Indonesia, social media is a 'hunting ground' for religious minorities. Rest of World Reporting Global Tech Stories, 23 July. <https://restofworld.org/2024/indonesia-muslim-influencers>
- Repnikova, M. (2018). China's lessons for fighting fake news. *Foreign Policy*, 6 September. <https://foreignpolicy.com/2018/09/06/chinas-lessons-for-fighting-fake-news>
- Repucci, S. & Slipowitz, A. (2022). *The Global Expansion of Authoritarian Rule*. Freedom House. <https://freedomhouse.org/report/freedom-world/2022/global-expansion-authoritarian-rule>
- Repucci, S. & Slipowitz, A. House. (2022). *Freedom in the World 2022: The Global Expansion of Authoritarian Rule*. Freedom House <https://freedomhouse.org/report/freedom-world/2022/global-expansion-authoritarian-rule>
- Reuters. (2019). China seeks to root out fake news and deepfakes with new online content rules. 29 November. www.reuters.com/article/idUSKBN1Y3OVT
- Rexhepi, R. (2023). Content moderation: How the EU and the US approach is striking a balance between protecting free speech and protecting public interest. *Trento Student Law Review*, 5(1), 67–97. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4433201
- Reyna, A. (2024). DMA and DSA effective enforcement – Key to success. *Journal of Antitrust Enforcement*, 12(2), 320–324. <https://doi.org/10.1093/jaenfo/jnae018>
- RIA (Research ICT Africa). (2023a). *Data for Democracy Action: Countering Disinformation in African Elections through Open Access to Data and Journalism as a Public Good*. Workshop Outcome Report. Research ICT Africa, South Africa.
- RIA. (2023b). *Information Integrity on Digital Platforms*. RIA input on United Nations Policy Brief 8, led by G. Berger.
- Ricaurte, P. (2022). Ethics for the majority world: AI and the question of violence at scale. *Media, Culture & Society*, 44(4), 726–745. <https://doi.org/10.1177/016344437221099612>

- Richards, N., & Hartzog, W. (2019). The pathologies of digital consent. *Washington University Law Review*, 96, 1461–1503. https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=6460&context=law_lawreview
- Richter, A. (2019). *Disinformation in the Media under Russian Law*. European Audiovisual Observatory.
- Richter, A. (2024). International legal responses to ‘propaganda for war’ in modern warfare. *Journal of International Media & Entertainment Law*, 10(1), 54–80. www.swlaw.edu/curriculum/law-review-journals/journal-international-media-entertainment-law
- Rieder, B., & Sire, G. (2014). Conflicts of interest and incentives to bias: A microeconomic critique of Google’s tangled position on the Web. *New Media & Society*, 16(2), 195–211. <https://doi.org/10.1177/1461444813481195>
- Rieder, B., & Skop, Y. (2021). The fabrics of machine moderation: Studying the technical, normative, and organizational structure of Perspective API. *Big Data & Society*, 8(2), 1–16. <https://doi.org/10.1177/20539517211046181>
- Riedl, M. J., Whipple, K. N., & Wallace, R. (2022). Antecedents of support for social media content moderation and platform regulation: The role of presumed effects on self and others. *Information, Communication & Society*, 25(11), 1632–1649. <https://doi.org/10.1080/1369118X.2021.1874040>
- Ritzer, G., Ryan, J. M., Hayes, S., Elliot, M., & Jandrić, P. (2024). McDonaldization and artificial intelligence. *Postdigital Science and Education*, 1–14. <https://doi.org/10.1007/s42438-024-00475-3>
- Robb, G., & Hawthorne, R. (2019). Net neutrality and market power: The case of South Africa. *Telecommunications Policy*, 43(9), 1–15. <https://doi.org/10.1016/j.telpol.2019.03.003>
- Robbins, J. W. (2020). Countering Russian disinformation. CSIS (Center for Strategic and International Studies) Blog post, 23 September. www.csis.org/blogs/post-soviet-post/countering-russian-disinformation
- Roberts, H., Hine, E., Taddeo, M., & Floridi, L. (2024). Global AI governance: Barriers and pathways forward. *International Affairs*, 100(3), 1275–1286. <https://doi.org/10.1093/ia/iaae073>
- Roberts, T., & Ali, A. M. (2021). *Opening and Closing Online Civic Space in Africa: An Introduction to the Ten Digital Rights Landscape Reports*. Institute for Development Studies, Sussex.
- Roberts, T., & Bosch, T. (2023a). Introduction: Spaces of Digital Citizenship in Africa. In T. Roberts & T. Bosch (Eds), *Digital Citizenship in Africa: Technologies of Agency and Repression* (pp. 1–32). Zed Books.
- Roberts, T., & Bosch, T. (Eds) (2023b). *Digital Citizenship in Africa: Technologies of Agency and Repression*. Zed Books.
- Roberts, T., Gitahi, J., Allam, P., Oboh, L., et al. (2023). *Mapping the Supply of Surveillance Technologies to Africa: Case Studies from Nigeria, Ghana, Morocco, Malawi, and Zambia*. Institute for Development Studies, Sussex.
- Robertson, C. E., del Rosario, K. S., & Van Bavel, J. J. (2024). Inside the funhouse mirror factory: How social media distorts perceptions of norms. *Current Opinion in Psychology*, 60, 1–7. <https://doi.org/10.1016/j.copsyc.2024.101918>
- Robinson, L., Schulz, J., Dodel, M., Correa, T., et al. (2020). Digital inclusion across the Americas and Caribbean. *Social Inclusion*, 8(2), 244–259. <https://doi.org/10.17645/si.v8i2.2632>
- Roche, C., Wall, P. J., & Lewis, D. (2023). Ethics and diversity in artificial intelligence policies, strategies and initiatives. *AI and Ethics*, 3(4), 1095–1115. <https://doi.org/10.1007/s43681-022-00218-9>
- Rodríguez-Ferrándiz, R. (2023). An overview of the fake news phenomenon: From untruth-driven to post-truth-driven approaches. *Media and Communication*, 11(2), 15–29. www.cogitatiopress.com/mediaandcommunication/article/view/6315
- Rogers, R. (2018). Digital traces in context| Otherwise engaged: Social media from vanity metrics to critical analytics. *International Journal of Communication*, 12(2018), 23. <https://ijoc.org/index.php/ijoc/article/view/6407>
- Rogers, R. (2020). The scale of Facebook’s problem depends upon how ‘fake news’ is classified. *Harvard Kennedy School Misinformation Review*, 1(6), 1–14. <https://doi.org/10.37016/mr-2020-43>
- Rohn, U., Bjørn von Rimscha, M., & Raats, T. (Eds) (2024). *De Gruyter Handbook of Media Economics*. De Gruyter.
- Romano, A. (2013). *Politics and the Press in Indonesia: Understanding an Evolving Political Culture*. Routledge.
- Romeu, A. (2023). New digital law tightens clampdown on press freedom in Cuba. Reporters Without Borders. <https://rsf.org/en/new-digital-law-tightens-clampdown-press-freedom-cuba>
- Ronzhyn, A., Cardenal, A. S., & Batlle Rubio, A. (2023). Defining affordances in social media research: A literature review. *New Media & Society*, 25(11), 3165–3188. <https://doi.org/10.1177/14614448221135187>
- Rosenberg, D. (2013). Data Before the Fact. In L. Gitelman (Ed.), *‘Raw Data’ is an Oxymoron* (pp. 15–40). The MIT Press.
- Rosol, M., & Blue, G. (2022). From the smart city to urban justice in a digital age. *City*, 26(4), 684–705. <https://doi.org/10.1080/13604813.2022.2079881>
- Ross Arguedas, A., & Simon, F. M. (2023). *Automating Democracy: Generative AI, Journalism, and the Future of Democracy*. Balliol College, Oxford Internet Institute and Institute for Ethics in AI.
- Ross Arguedas, A., Robertson, C., Fletcher, R., & Nielsen, R. (2022a). *Echo Chambers, Filter Bubbles, and Polarisation: A Literature Review*. Reuters Institute for the Study of Journalism, University of Oxford and The Royal Society.
- Ross Arguedas, A., Badrinathan, S., Mont’Alverne, C., Toff, B., Fletcher, R., & Nielsen, R. K. (2022b). ‘It’s a battle you are never going to win’: Perspectives from journalists in four countries on how digital media platforms undermine trust in news. *Journalism Studies*, 23(14), 1821–1840. <https://doi.org/10.1080/1461670X.2022.2112908>
- Ross Arguedas, A., Badrinathan, S., Mont’Alverne, C., Toff, B., Fletcher, R., & Nielsen, R. (2022c). *Snap Judgements: How Audiences Who Lack Trust in News Navigate Information on Digital Platforms*. Reuters Institute for the Study of Journalism, University of Oxford.

- Roth, E. (2022). Google results for abortion clinics are misleading and politically fraught. *The Verge*, 16 August. www.theverge.com/2022/8/16/23307850/google-maps-results-abortion-clinics-crisis-pregnancy-centers
- Rouvroy, A., & Berns, T. (2013). Algorithmic governmentality and prospects of emancipation. *Rezeaux*, 177(1), 163–196. www.cairn-int.info/article-E_RES_177_0163--algorithmic-governmentality-and-prospect.htm?contenu=plan
- Roy, A. (2004). Event summary: Poverty and the right to know. Brookings Institution Commentary, 28 October. www.brookings.edu/articles/event-summary-poverty-and-the-right-to-know
- Roy, R. (2024). Big relief! Supreme Court stays notification constituting Fact-Check Unit! Internet Freedom Foundation, 23 March. <https://internetfreedom.in/sc-stays-notification-constituting-fcu>
- RSF (Reporters sans frontières [Reporters Without Borders]). (2022). Methodology used for compiling World Press Freedom Index. <https://rsf.org/en/index-methodologie-2022>
- RSF. (2024). 2024 World Press Freedom Index – journalism under political pressure. <https://rsf.org/en/2024-world-press-freedom-index-journalism-under-political-pressure>
- Ruffo, G., Semeraro, A., Giachanou, A., & Rosso, P. (2023). Studying fake news spreading, polarisation dynamics, and manipulation by bots: A tale of networks and language. *Computer Science Review*, 47(2023), 1–26. <https://doi.org/10.1016/j.cosrev.2022.100531>
- Ruggie, J. G. (2011). *Report of the Special Representative of the Secretary-General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises: Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework*. United Nations General Assembly A/HRC/17/31. <https://digitallibrary.un.org/record/705860?v=pdf>
- Rugh, W. A. (2000). *Arab Mass Media: Newspapers, Radio, and Television in Arab Politics*. Praeger Publishers.
- Ruohonen, J. (2003). David and Goliath: Privacy Lobbying in the European Union. In J. Greenwood (Ed.), *The Challenge of Change in EU Business Associations* (pp. 139–149). Palgrave Macmillan.
- Rusbridger, A. (2024). Elon Musk's misinformation machine made the horrors of Southport much worse. *Prospect*, 2 August. www.prospectmagazine.co.uk/ideas/media/67463/elon-musks-misinformation-machine-made-the-horrors-of-southport-much-worse
- Ryan-Mosley, T. (2023). How generative AI is boosting the spread of disinformation and propaganda. MIT Technology Review, 4 October. www.technologyreview.com/2023/10/04/1080801/generative-ai-boosting-disinformation-and-propaganda-freedom-house/
- Ryng, J., Guicherd, G., Saman, J. A., Choudhury, P., & Kellett, A. (2022). Internet shutdowns: A human rights issue. *The RUSI Journal*, 167(4–5), 50–63. <https://doi.org/10.1080/03071847.2022.2156234>
- Sacasas, L. M. (2021). The materiality of digital culture: It's not as invisible – or as disembodiment – as our devices lead us to believe. *Comment*, 12 August. <https://comment.org/the-materiality-of-digital-culture>
- Sádaba, C., Salaverría, R., & Bingué, X. (2023). Overcoming the age barrier: Improving older adults' detection of political disinformation with media literacy. *Media and Communication*, 11(4), 113–123. <https://doi.org/http://dx.doi.org/10.17645/mac.v11i4.7090>
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1), 1–12. <https://doi.org/10.1177/2053951718820549>
- Salehi, N., Irani, L. C., Bernstein, M. S., Alkhatib, A., Ogbe, E., Milland, K., & Clickhappier. (2015). We are dynamo: Overcoming stalling and friction in collective action for crowd workers. *CHI '15*, 18–23 April. <http://dx.doi.org/10.1145/2702123.2702508>
- Samoilenko, S. A., & Suvorova, I. (2023). Artificial Intelligence and Deepfakes in Strategic Deception Campaigns: The US and Russian Experiences. In E. Pashentsev (Ed.), *The Palgrave Handbook of Malicious Use of AI and Psychological Security* (pp. 507–529). Springer International Publishing.
- Samoli, S., Lopez, C. M., Gomez, G. E., De, P. G., Martinez-Plumed, F., & Delipetrev, B. (2020). *AI WATCH. Defining Artificial Intelligence*. European Commission, JRC EUR30117.
- Sampath, P. G. (2021). Governing artificial intelligence in an age of inequality. *Global Policy*, 12(S6), 21–31. <https://doi.org/10.1111/1758-5899.12940>
- Sančanin, B., & Penjišević, A. (2022). Use of artificial intelligence for the generation of media content. *Social Informatics Journal*, 1(1), 1–7. <https://doi.org/10.58898/sij.v1i1.01-07>
- Sanders, N. E., & Schneider, B. (2024). Let's not make the same mistakes with AI that we made with social media. *MIT Technology Review*, 13 March. www.technologyreview.com/2024/03/13/1089729/lets-not-make-the-same-mistakes-with-ai-that-we-made-with-social-media
- Sankaran, V. (2024). Elon Musk's X takes on Indian government in freedom of speech row. *The Independent*, 22 February. www.independent.co.uk/tech/x-twitter-accounts-withheld-india-modi-b2500486.html
- Santini, R. M., Agostini, L., Barros, C. E., Carvalho, D., et al. (2018). Software power as soft power. A literature review on computational propaganda effects in public opinion and political process. *The Open Journal of Sociopolitical Studies*, 11(2), 1–29. <https://doi.org/10.1285/i20356609v11i2p332>
- Santini, R. M., Salles, D., & Barros, C. E. (2022). We love to hate George Soros: A cross-platform analysis of the Globalism conspiracy theory campaign in Brazil. *Convergence: The International Journal of Research into New Media Technologies*, 28(4), 983–1006. <https://doi.org/10.1177/13548565221085833>

- Santini, R. M., Salles, D., Borges, M., Medeiros, P., et al. (2023). *The Permanent Campaign Against the Press in Brazil: Analysis of the Media Ecosystem and Multiplatform Digital Advertising*. NetLab – Laboratory of Internet and Social Network Studies, Federal University of Rio de Janeiro (UFRJ).
- Santini, R. M., Salles, D., Tucci, G., Ferreira, F., & Graef, F. (2020). Making up audience: Media bots and the falsification of the public sphere. *Communication Studies*, 71(3), 466–487. <https://doi.org/10.1080/10510974.2020.1735466>
- Santos, P., & Ndlovu, K. (2022). Towards a journalism for justice: A normative overture. *African Journalism Studies*, 43(1), 70–88. <https://doi.org/10.1080/23743670.2021.2009533>
- Saurwein, F., & Spencer-Smith, C. (2020). Combating disinformation on social media: Multilevel governance and distributed accountability in Europe. *Digital Journalism*, 8(6), 820–841. <https://doi.org/10.1080/21670811.2020.1765401>
- Saurwein, F., & Spencer-Smith, C. (2020). Combating disinformation on social media: Multilevel governance and distributed accountability in Europe. *Digital Journalism*, 8(6), 820–841. <https://doi.org/10.1080/21670811.2020.1765401>
- Savitz, E. J. (2019). Google buys data-analytics firm Looker amid privacy concerns. *Baron's*, 6 June. www.barrons.com/articles/google-buys-data-analytics-firm-looker-privacy-concerns-51559837661
- Schaaake, M., & Fukuyama, F. (Eds) (2023). *Digital Technologies in Emerging Countries*. Stanford Cyber Policy Center.
- Schaetz, N., Gagrčin, E., Toth, R., & Emmer, M. (2023). Algorithm dependency in platformized news use. *New Media & Society, Online First*, 1–18. <https://doi.org/10.1177/14614448231193093>
- Schippers, B. (2020). Artificial intelligence and democratic politics. *Political Insight*, 11(1), 32–35. <https://journals.sagepub.com/doi/full/10.1177/2041905820911746>
- Schirch, L. (2021). Digital Information, Conflict and Democracy. In L. Schirch (Ed.), *Social Media Impacts on Conflict and Democracy*. Routledge.
- Schlesinger, P. (2020). After the post-public sphere. *Media, Culture & Society*, 42(7–8), 1545–1563. <https://doi.org/10.1177/0163443720948003>
- Schlosberg, J., & Freedman, D. (2020). Opening the gates: Plurality regulation in a post digital world. *Journal of Digital Media & Policy*, 11(2), 115–132. https://intellectdiscover.com/content/journals/10.1386/jdmp_00017_1
- Schmarzo, B. (2023). *AI & Data Literacy: Empowering Citizens of Data Science*. Packt Publishing Ltd.
- Schmitt, V., Villa-Arenas, L. F., Feldhus, N., Meyer, J., et al. (2024). The role of explainability in collaborative human-AI disinformation detection. *FACCT'24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. <https://doi.org/10.1145/3630106.3659031>
- Schneider, N. (2022). Governable spaces: A feminist agenda for platform policy. *Internet Policy Review*, 11(1), 1–19. <https://doi.org/10.14763/2022.1.1628>
- Schoon, A., Mabweazara, H. M., Bosch, T., & Dugmore, H. (2020). Decolonising digital media research methods: Positioning African digital experiences as epistemic sites of knowledge production. *African Journalism Studies*, 41(4), 1–15. <https://doi.org/10.1080/23743670.2020.1865645>
- Schot, E. (2020). *Income Models of Independent Media in Difficult Contexts*. Free Press Unlimited.
- Schranz, M., Schneider, J., & Eisenegger, M. (2018). Media Trust and Media Use. In K. Otto & A. Köhler (Eds), *Trust in Media and Journalism: Empirical Perspectives on Ethics, Norms, Impacts and Populism in Europe* (pp. 73–91). Springer Fachmedien.
- Schreurs, K., Quan-Haase, A., & Martin, K. (2017). Problematizing the digital literacy paradox in the context of older adults' ICT use: Aging, media discourse, and self-determination. *Canadian Journal of Communication*, 42(2), 359–377. <https://doi.org/http://doi.org/10.22230/cjc.2017v42n2a3130>
- Schudson, M. (1978). *Discovering the News: A Social History of American Newspapers*. Basic Books.
- Schudson, M. (2022). What does 'trust in the media' mean? *Daedalus*, 151(4), 144–160. https://doi.org/10.1162/daed_a_01948
- Schünemann, W. J. (2022). A Threat to Democracies? An Overview of Theoretical Approaches and Empirical Measurements for Studying the Effects of Disinformation. In M. D. Cavelty & A. Wenger (Eds), *Cyber Security Politics* (pp. 32–47). Routledge.
- Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, US National Institute of Standards and Technology.
- Scott, M., Bunce, M., & Wright, K. (2022). The influence of news coverage on humanitarian aid: The bureaucrats' perspective. *Journalism Studies*, 23(2), 167–186. <https://doi.org/10.1080/1461670X.2021.2013129>
- Segura, M. S., & Waisbord, S. (2019). Between data capitalism and data citizenship. *Television & New Media*, 20(4), 412–419. <https://doi.org/10.1177/1527476419834519>
- Seipp, T. J., Fathaigh, R. Ó., & van Drunen, M. (2023a). Defining the 'media' in Europe: Pitfalls of the proposed European Media Freedom Act. *Journal of Media Law*, 15(1), 39–51. <https://doi.org/10.1080/17577632.2023.2240998>
- Seipp, T. J., Helberger, N., de Vreese, C., & Ausloos, J. (2023b). Dealing with opinion power in the platform world: Why we really have to rethink media concentration law. *Digital Journalism*, 11(8), 1542–1567. <https://doi.org/10.1080/21670811.2022.2161924>
- Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, 2019/01/29/). *Fairness and Abstraction in Sociotechnical Systems*. FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency, 59–68. <https://doi.org/10.1145/3287560.3287598>

- Sepas, A., Bangash, A. H., Alraoui, O., El Emam, K., & El-Hussuna, A. (2022). Algorithms to anonymize structured medical and healthcare data: A systematic review. *Frontiers in Bioinformatics*, 2, 1-11. <https://doi.org/10.3389/fbinf.2022.984807>
- Settle, J. E. (2018). *Frenemies: How Social Media Polarizes America*. Cambridge University Press.
- Sey, A., Shilongo, K., Hlomani, H., & Diouf, D. (2022). Sub-Saharan Africa. In H. Wasserman (Ed.), *Meeting the Challenges of Information Disorder in the Global South* (pp. 138-189). IDRC Research Results.
- Shahbaz, A., Funk, A., & Vesteinsson, K. (2022). *Countering an Authoritarian Overhaul of the Internet*. Freedom House.
- Shahin, S. (2019). Facing up to Facebook: How digital activism, independent regulation, and mass media foiled a neoliberal threat to net neutrality. *Information, Communication & Society*, 22(1), 1-17. <https://doi.org/10.1080/1369118X.2017.1340494>
- Shapero, J. (2024). Musk calls Brazilian judge 'evil dictator' amid X feud. *The Hill*, 30 August. <https://thehill.com/policy/technology/4855584-musk-slams-brazilian-judge-threatens-x/>
- Shapiro, R. (2019). Transnational networks of literacy and materiality: Coltan, sexual violence, and digital literacy. *College English*, 82(2), 204-225. <https://doi.org/10.58680/ce201930626>
- Sheombar, A., & Skelton, S. K. (2023). Follow the Surveillance: A Breadcrumb Trail of Surveillance Technology Exports to Africa. In M. R. Jones, A. S. Mukherjee, D. Thapa, & Y. Zheng (Eds), *After Latour: Globalisation, Inequity and Climate Change* (pp. 241-261). Springer Nature.
- Sherman, J. (2022). India's sudden reversal on privacy will affect the global internet. *Slate*, 5 September. <https://slate.com/technology/2022/09/india-data-protection-bill-fourth-way.html>
- Shin, D.-H., & Lee, M.-K. (2017). Public value mapping of network neutrality: Public values and net neutrality in Korea. *Telecommunications Policy*, 41(3), 208-224. <https://doi.org/10.1016/j.telpol.2016.12.012>
- Shin, D.-H., Kee, K. F., & Shin, E. Y. (2022). Algorithm awareness: Why user awareness is critical for personal privacy in the adoption of algorithmic platforms? *International Journal of Information Management*, 65(2022), 1-13. <https://doi.org/10.1016/j.ijinfomgt.2022.102494>
- Shirikov, A. (2024). Fake news for all: How citizens discern disinformation in autocracies. *Political Communication*, 41(1), 45-65. <https://doi.org/10.1080/10584609.2023.2257618>
- Shleifer, A. (2004). *Does Competition Destroy Ethical Behavior?* NBER Working Paper 10269. National Bureau of Economic Research. www.nber.org/papers/w10269
- Shoemaker, P. J., & Vos, T. (2009). *Gatekeeping Theory*. Routledge.
- Siapera, E. (2023). The Challenge of Far-Right Media. In S. Farrell, E. Siapera, & G. Souvlis (Eds), *Radical Journalism: Resurgence, Reform, Reaction* (Chapter 9). Routledge.
- Silverstone, R. (2007). *Media and Morality: On the Rise of the Mediapolis*. Polity Press.
- Simon, F. M. (2022). Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism*, 10(10), 1832-1854. <https://doi.org/10.1080/21670811.2022.2063150>
- Simon, F. M. (2024). *Artificial Intelligence in the News: How AI Retools, Rationalizes, and Reshapes Journalism and the Public Arena*. Tow Center for Journalism, Oxford Internet Institute, University of Oxford.
- Simon, F. M., & Isaza-Ibarra, L. F. (2023). *AI in the News: Reshaping the Information Ecosystem?* Balliol College, Torch, Reuters Institute, University of Oxford.
- Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School (HKS) Misinformation Review*, 4(5), 1-11. <https://doi.org/10.37016/mr-2020-127>
- Singapore Statutes. (2019). *Protection from Online Falsehoods and Manipulation Act 2019*. Government of Singapore.
- Singh, P. J. (2020). *Economic Rights in a Data-Based Society: Collective Data Ownership, Workers' Rights, and the Role of the Public Sector*. Public Services International, Friedrich Ebert Foundation and IT for Change, India.
- Singh, P. J., & Gurumurthy, A. (2021). *Economic Governance of Data: Balancing Individualist-Property Approaches with a Community Rights Framework*. IT for Change, India.
- Sinpeng, A. (2020). Digital media, political authoritarianism, and internet controls in Southeast Asia. *Media, Culture & Society*, 42(1), 25-39. <https://doi.org/10.1177/0163443719884052>
- Sirbu, A., Pedreschi, D., Giannotti, F., & Kertész, J. (2019). Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. *PLOS ONE*, 14(3), 1-20. <https://doi.org/10.1371/journal.pone.0213246>
- Skalli, L. H. (2023). Disruptive Social Entrepreneurship from Bahrain: The Case of Esra'a Al Shafei. In L. H. Skalli & N. Eltantawy (Eds), *The Palgrave Handbook of Gender, Media and Communication in the Middle East and North Africa* (pp. 567-583). Springer International Publishing.
- Skibinski, M. (2021). Special Report: Top brands are sending \$2.6 billion to misinformation websites each year. *NewsGuard*. www.newsguardtech.com/special-reports/brands-send-billions-to-misinformation-websites-newsguard-comscore-report
- Skovsgaard, M., & Andersen, K. (2020). Conceptualizing news avoidance: Towards a shared understanding of different causes and potential solutions. *Journalism Studies*, 21(4), 459-476. <https://doi.org/10.1080/1461670X.2019.1686410>
- Sleibi, A. (2023). Wartime Opportunities: The Sway of Russian Disinformation on the MENA Region. In EIMed (European Institute of the Mediterranean) (Ed.), *Wartime-Russia Yearbook 2023* (pp. 312-317). www.iemed.org/med-yearbook/iemed-mediterranean-yearbook-2023

- Smith, J. R. (2019). IBM Research releases 'Diversity in Faces' dataset to advance study of fairness in facial recognition systems. *Phys.org*, 29 January. <https://phys.org/news/2019-01-ibm-diversity-dataset-advance-fairness.html>
- Söderström, O., & Datta, A. (Eds) (2023). *Data Power in Action: Urban Data Politics in Times of Crisis*. Bristol University Press.
- Sodré, M. (2021). *A sociedade incivil: Mídia, iliberalismo e finanças*. Editora Vozes.
- Solomun, S., Polataiko, M., & Hayes, H. A. (2021). Platform responsibility and regulation in Canada: Considerations on transparency, legislative clarity, and design. *Harvard Journal of Law & Technology*, 34, 1–18. <https://jolt.law.harvard.edu/digest/platform-responsibility-and-regulation-in-canada-considerations-on-transparency-legislative-clarity-and-design>
- Souter, D. (2024). Inside the digital society: Governing AI. APC (Association for Progressive Communications) Talk, 1 October. www.apc.org/en/blog/inside-digital-society-governing-ai
- South Africa Government. (2024). *National Policy on Data and Cloud*. Communications & Digital Technologies Department, Republic of South Africa.
- Spitale, G., Biller-Andorno, N., & Germani, F. (2023). AI model GPT-3 (dis)informs us better than humans. *Science Advances*, 9(26), 1–9. <https://doi.org/10.1126/sciadv.adh1850>
- Splichal, S. (2022a). *Datafication of Public Opinion and the Public Sphere: How Extraction Replaced Expression of Opinion*. Anthem Press.
- Splichal, S. (2022b). The public sphere in the twilight zone of publicness. *European Journal of Communication*, 37(2), 198–215. <https://doi.org/10.1177/02673231211061490>
- SSD EFF (Surveillance Self-Defense, Electronic Frontier Foundation) (2023). Surveillance Self-Defense. <https://ssd.eff.org>
- Stamboliev, E. (2023). Proposing a postcritical AI literacy: Why we should worry less about algorithmic transparency and more about citizen empowerment. *Media Theory*, 7(1), 202–232. <https://hal.science/hal-04221538>
- Stanly, M. (2024). The future of work: Ensuring AI literacy for all. INDIAai, 4 July. <https://indiaai.gov.in/article/the-future-of-work-ensuring-ai-literacy-for-all>
- Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, 7(1), 111–134. www.jstor.org/stable/23010792
- Steensen, S., Grøndahl Larsen, A. M., Hågvær, Y. B., & Fonn, B. K. (2019). What does digital journalism studies look like? *Digital Journalism*, 7(3), 320–342. <https://doi.org/10.1080/21670811.2019.1581071>
- Stefanija, A. P., & Pierson, J. (2023). Algorithmic governmentality, digital sovereignty, and agency affordances: Extending the possible fields of action. *Weizenbaum Journal of the Digital Society*, 3(2), 1–30. <https://doi.org/10.34669/WI.WJDS/3.2.2>
- Štětka, V., & Mihelj, S. (2024a). The Rise of the Illiberal Public Sphere. In V. Štětka & S. Mihelj (Eds), *The Illiberal Public Sphere: Media in Polarized Societies* (pp. 25–59). Springer Nature Switzerland
- Štětka, V., & Mihelj, S. (2024b). *The Illiberal Public Sphere: Media in Polarized Societies*. Springer Nature Switzerland.
- Stevenson, C. (2016). Seattle Surveillance Ordinance Memo. Technology Law and Public Policy Clinic, 6 December. <https://digitalcommons.law.uw.edu/cgi/viewcontent.cgi?article=1002&context=techclinic>
- Stiefel, L., Currie, M., Musiani, F., Sandoz, A., Silvest, A., & Williams, R. (2024). Preface: Governance by infrastructure. *First Monday*, 29(2), 1–4. <https://firstmonday.org/ojs/index.php/fm/article/view/13559/11406>
- Stigler Committee. (2019). *Stigler Committee on Digital Platforms: Final Report*. Stigler Center for the Study of the Economy and the State, University of Chicago Booth School of Business.
- Stolton, S., & Maksimov, V. (2020). Orbán to rule by decree with new powers to 'silence critics'. Euractiv, 30 March. www.euractiv.com/section/global-europe/news/orban-to-rule-by-decree-with-new-powers-to-silence-critics
- Stray, J., Iyer, R., & Larrauri, H. P. (2023). *The Algorithmic Management of Polarization and Violence on Social Media*. Knight First Amendment Institute, Columbia University.
- Strömbäck, J., & Shehata, A. (2019). The reciprocal effects between political interest and TV news revisited: Evidence from four panel surveys. *Journalism & Mass Communication Quarterly*, 96(2), 473–496. <https://doi.org/10.1177/1077699018793998>
- Strömbäck, J., Falasca, K., & Kruikemeier, S. (2018). The mix of media use matters: Investigating the effects of individual news repertoires on offline and online political participation. *Political Communication*, 35(3), 413–432. <https://doi.org/10.1080/10584609.2017.1385549>
- Strömbäck, J., Tsifti, Y., Boomgaarden, H., Damstra, A., et al. (2020). News media trust and its impact on media use: Toward a framework for future research. *Annals of the International Communication Association*, 44(2), 139–156. <https://doi.org/10.1080/08023808985.2020.1755338>
- Stroud, N. J. (2017). Selective exposure theories. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford Handbook of Political Communication* (pp. 531–548). Oxford University Press.
- Stroud, N. J., & Lee, J. K. (2013). Perceptions of cable news credibility. *Mass Communication and Society*, 16(1), 67–88. <https://doi.org/10.1080/15205436.2011.646449>
- Strover, S., Choi, J., & Schrubbe, A. (2024). Broadband, rural contexts and local economic dynamics. *Telecommunications Policy*, 48(3), 1–12. <https://doi.org/https://doi.org/10.1016/j.telpol.2024.102720>

- Suber, P. (2012). *Open Access*. The MIT Press.
- Suchman, L. (2023). The uncontroversial 'thingness' of AI. *Big Data & Society*, 10(2), 1–5. <https://doi.org/10.1177/20539517231206794>
- Suiter, J., & Fletcher, R. (2020). Polarization and partisanship: Key drivers of distrust in media old and new? *European Journal of Communication*, 35(5), 484–501. <https://doi.org/10.1177/0267323120903685>
- Sultan, M., Tump, A. N., Ehmann, N., Lorenz-Spreen, P., et al. (2024). Susceptibility to online misinformation: A systematic meta-analysis of demographic and psychological factors. *PsyArXiv*. <https://doi.org/https://osf.io/preprints/psyarxiv/7u4fg>
- Sunstein, C. (2007). *Republic.com 2.0*. Princeton University Press.
- Szakacs, J., & Bogнар, É. (2021). *The Impact of Disinformation Campaigns about Migrants and Minority Groups in the EU*. European Parliament, Directorate General for External Policies of the Union.
- Szczepkowski, D., & Szczepkowski, P. (2021). Adware as one form of disinformation. *Studia Administracji i Bezpieczeństwa*, 1(10), 197–208. www.ceeol.com/search/article-detail?id=1004222
- Talamanca, G., & Arfini, S. (2022). Through the newsfeed glass: Rethinking filter bubbles and echo chambers. *Philosophy and Technology*, 35(1), 1–34. <https://doi.org/doi:10.1007/s13347-021-00494-z>
- Tambini, D. (2021). *Media Freedom*. Polity.
- Tambini, D. (2021). What is journalism? The paradox of media privilege. *European Human Rights Law Review*, 5, 523–539. https://eprints.lse.ac.uk/112681/1/What_is_journalism_The_paradox_of_media_privilege_clean.pdf
- Tan, C. (2020). *Regulating Content on Social Media: Copyright, Terms of Service and Technological Features*. Creative Media Partners, LLC.
- Tandoc Jr, E. C., & Vos, T. P. (2016). The journalist is marketing the news: Social media in the gatekeeping process. *Journalism Practice*, 10(8), 950–966. <https://doi.org/10.1080/17512786.2015.1087811>
- Taneja, H., Webster, J. G., Malthouse, E. C., & Ksiazek, T. B. (2012). Media consumption across platforms: Identifying user-defined repertoires. *New Media & Society*, 14(6), 951–968. <https://doi.org/10.1177/1461444811436146>
- Tang, S., Willnat, L., & Zhang, H. (2021). Fake news, information overload, and the third-person effect in China. *Global Media and China*, 6(4), 492–507. <https://doi.org/10.1177/20594364211047369>
- Tapper, J. (2024). The far right has moved online, where its voice is more dangerous than ever. *The Observer*, 3 August. www.theguardian.com/politics/article/2024/aug/03/the-far-right-has-moved-online-where-its-voice-is-more-dangerous-than-ever
- Tarrant, A., & Cowen, T. (2022). Big tech lobbying in the EU. *The Political Quarterly*, 93(2), 218–226. <https://doi.org/10.1111/1467-923X.13127>
- Tate-Mosely, R. (2023). The movement to limit face recognition tech might finally get a win. *MIT Technology Review*, 20 June. www.technologyreview.com/2023/07/20/1076539/face-recognition-massachusetts-test-police
- Tay, L. Q., Lewandowsky, S., Hurlstone, M. J., Kurz, T., & Ecker, U. K. H. (2024). Thinking clearly about misinformation. *Communications Psychology*, 2(1), 1–5. <https://doi.org/10.1038/s44271-023-00054-5>
- Taylor, L., Mukiri-Smith, H., Petročnik, T., Savolainen, L., & Martin, A. (2022). (Re)making data markets: An exploration of the regulatory challenges. *Law, Innovation and Technology*, 14(2), 355–394. <https://doi.org/10.1080/17579961.2022.2113671>
- Teachout, Z., & Khan, L. (2014). Market structure and political law: A taxonomy of power. *Duke Journal of Constitutional Law & Public Policy*, 9, 37–74. <https://doi.org/10.2139/ssrn.2490525>
- Tech Hive Advisory Center for Law & Innovation. (2024). *State of AI Regulation in Africa: Trends and Developments Report*. www.techhiveadvisory.africa/report/state-of-ai-regulation-in-africa-trends-and-developments
- Teixeira da Silva, J. A. (2021). How to shape academic freedom in the digital age? Are the retractions of opinionated papers a prelude to 'cancel culture' in academia? *Current Research in Behavioral Sciences*, 2(2021), 1–6. <https://doi.org/10.1016/j.crbeha.2021.100035>
- Tenove, C. (2020). Protecting democracy from disinformation: Normative threats and policy responses. *The International Journal of Press/Politics*, 25(3), 517–537. <https://doi.org/10.1177/1940161220918740>
- Teyssou, D., Leung, J.-M., Apostolidis, E., Apostolidis, K., et al. (2017). The InVID Plug-in: Web Video Verification on the Browser. In ACM (Association for Computing Machinery) (Ed.), *Proceedings of the First International Workshop on Multimedia Verification* (pp. 23–30). <https://doi.org/10.1145/3132384.3132387>
- Thakur, D., & Madrigal, D. H. (2022). *An Unrepresentative Democracy: How Disinformation and Online Abuse Hinder Women of Color Political Candidates in the United States*. Center for Democracy & Technology, US & Belgium.
- The Majority Report w/ Sam Seder. (2023). Data cartels and our data. www.youtube.com/watch?v=E-VWLIFUXIA
- The Tierra Común Network (Ed.) (2023). *Resisting Data Colonialism: A Practical Intervention*. Institute of Network Cultures, Amsterdam University. https://networkcultures.org/wp-content/uploads/2023/12/ResistingDataColonialism_INC2023_TOD50.pdf
- Thomas, E. (2022). *Conspiracy Clickbait: This One Weird Trick Will Undermine Democracy*. ISD Institute for Strategic Studies.
- Thompson, A., & Kemp, S. (2024). Digital 2024: 5 billion social media users. We Are Social, 31 January. <https://wearesocial.com/uk/blog/2024/01/digital-2024-5-billion-social-media-users>

- Thussu, D. K. (2022). De-colonizing global news-flows: A historical perspective. *Journalism Studies*, 23(13), 1578–1592. <https://doi.org/10.1080/1461670X.2022.2083007>
- Timcke, S., & Hlomani, H. (2024). *How Might AI Reshape Democracy on the African Continent?* Research ICT Africa, South Africa.
- Toff, B., & Kalogeropoulos, A. (2020). All the news that's fit to ignore: How the information environment does and does not shape news avoidance. *Public Opinion Quarterly*, 84(S1), 366–390. <https://doi.org/10.1093/poq/nfaa016>
- Toff, B., & Nielsen, R. K. (2018). 'I just Google it': Folk theories of distributed discovery. *Journal of Communication*, 68(3), 636–657. <https://doi.org/10.1093/joc/jqy009>
- Toff, B., & Palmer, R. (2019). Explaining the gender gap in news avoidance: 'News-is-for-men' perceptions and the burdens of caretaking. *Journalism Studies*, 20(11), 1563–1579. <https://doi.org/10.1080/1461670X.2018.1528882>
- Toff, B., Badrinathan, S., Mont'Alverne, C., Ross Arguedas, A., Fletcher, R., & Nielsen, R. (2021a). *Overcoming Indifference: What Attitudes Towards News Tell Us about Building Trust*. Reuters Institute for the Study of Journalism, University of Oxford.
- Toff, B., Badrinathan, S., Mont'Alverne, C., Ross Arguedas, A., Fletcher, R., & Nielsen, R. (2021b). *Listening to What Trust in News Means to Users: Qualitative Evidence from Four Countries*. Reuters Institute for the Study of Journalism, University of Oxford.
- Toff, B., Palmer, R., & Nielsen, R. K. (2023). *Avoiding the News: Reluctant Audiences for Journalism*. Columbia University Press.
- Tollefson, J. (2024). 'Vindicated': Embattled misinformation researchers celebrate key US Supreme Court decision. *Nature*, 26 June. www.nature.com/articles/d41586-024-01766-2
- Tomaz, T., & Trappel, J. (2022). News media in European democracies and beyond: Stable structural conditions but notorious deficits. *De Europa*, 5(2), 17–34. <https://doi.org/10.13135/2611-853X/7200>
- Tong, C., Gill, H., Li, J., Valenzuela, S., & Rojas, H. (2020). 'Fake news is anything they say!' – Conceptualization and weaponization of fake news among the American public. *Mass Communication & Society*, 23(5), 755–778. <https://doi.org/10.1080/15205436.2020.1789661>
- Tonnies, F. (1957). *Community & Society [Gemeinschaft und Gesellschaft]*, C. P. Loomis, translator. Harper & Row [first published 1887].
- Trappel, J. (Ed.) (2019). *Digital Media Inequalities: Policies against Divides, Distrust and Discrimination*. Nordicom.
- Trauthig, I. (2024). *Diaspora Communities and Computational Propaganda on Messaging Apps*. CIGI (Centre for International Governance Innovation), Canada.
- Tsanni, A. (2024). How satellite images and AI could help fight spatial apartheid in South Africa. *MIT Technology Review*, 19 January. www.technologyreview.com/2024/01/19/1086837/satellite-images-ai-spatial-apartheid-south-africa
- Tsfaty, Y., & Ariely, G. (2014). Individual and contextual correlates of trust in media across 44 countries. *Communication Research*, 41(6), 760–782. <https://doi.org/10.1177/0093650213485972>
- Tsfaty, Y., Boomgaarden, H. G., Strömbäck, J., Vliegenthart, R., Damstra, A., & Lindgren, E. (2020). Causes and consequences of mainstream media dissemination of fake news: Literature review and synthesis. *Annals of the International Communication Association*, 44(2), 157–173. <https://doi.org/10.1080/23808985.2020.1759443>
- Tuchman, G. (1972). Objectivity as strategic ritual: An examination of newsmen's notions of objectivity. *American Journal of Sociology*, 77(4), 660–679. <https://doi.org/10.1086/225193>
- Tucker, E. (2023). Our future inside the Fifth Column – Or, What chatbots are really for. *Tech Policy Press*, 14 June. www.techpolicy.press/our-future-inside-the-fifth-column-or-what-chatbots-are-really-for
- Tucker, J. A., Guess, A., Barbera, P., Vaccari, C., et al. (2018). *Social Media, Political Polarization, and Political Disinformation: A Review of the Literature*. Hewlett Foundation, US.
- Tully, M. (2022). Everyday news use and misinformation in Kenya. *Digital Journalism*, 10(1), 109–127. <https://doi.org/10.1080/21670811.2021.1912625>
- Tully, M., Madrid-Morales, D., Wasserman, H., Gondwe, G., & Ireri, K. (2022). Who is responsible for stopping the spread of misinformation? Examining audience perceptions of responsibilities and responses in six sub-Saharan African countries. *Digital Journalism*, 10(5), 679–697. <https://doi.org/10.1080/21670811.2021.1965491>
- Turillazzi, A., Taddeo, M., Floridi, L., & Casolari, F. (2023). The Digital Services Act: An analysis of its ethical, legal, and social implications. *Law, Innovation and Technology*, 15(1), 83–106. <https://doi.org/10.1080/17579961.2023.2184136>
- Turow, J., Lelkes, Y., Draper, N., & Waldman, A. E. (2023). *Americans Can't Consent to Companies' Use of Their Data: They Admit They Don't Understand It, Say They're Helpless to Control It, and Believe They're Harmed When Firms Use Their Data – Making What Companies Do Illegitimate*. Annenberg School for Communications, University of Pennsylvania. www.asc.upenn.edu/sites/default/files/2023-02/Americans_Cant_Consent.pdf
- Tusikov, N. (2021). Internet Platforms Weaponizing Chokepoints. In D. W. Drezner, H. Farrell, & A. Newman (Eds), *The Uses and Abuses of Weaponized Interdependence* (pp. 133–148). Brookings Institution Press.
- Tuters, M., & Hagen, S. (2020). (((They))) rule: Memetic antagonism and nebulous othering on 4chan. *New Media & Society*, 22(12), 2218–2237. <https://doi.org/10.1177/1461444819888746>

- Twomey, J., Ching, D., Aylett, M. P., Quayle, M., Linehan, C., & Murphy, G. (2023). Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *PLOS ONE*, 18(10), 1–22. <https://doi.org/10.1371/journal.pone.0291668>
- Tyagi, K., Sanders, A. K., & Cauffman, C. (Eds) (2024). *Digital Platforms, Competition Law, and Regulation: Comparative Perspectives*. Bloomsbury Publishing.
- UAE (United Arab Emirates). (2018). *UAE National Strategy for Artificial Intelligence*. National Program for Artificial Intelligence, <https://ai.gov.ae/wp-content/uploads/2021/07/UAE-National-Strategy-for-Artificial-Intelligence-2031.pdf>.
- UK DSIT (Department for Science, Innovation & Technology). (2023). *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023* UK Dept for Science, Innovation & Technology. www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023
- UK Government. (2003). *Communications Act 2003*. <https://www.legislation.gov.uk/ukpga/2003/21/contents>
- UK Government. (2023). *Online Safety Act 2023*. <https://www.legislation.gov.uk/ukpga/2023/50>
- Ulbricht, L., & Egbert, S. (2024). In Palantir we trust? Regulation of data analysis platforms in public security. *Big Data & Society*, 11(3), 1–15. <https://doi.org/10.1177/20539517241255108>
- UN (United Nations). (1948). *Universal Declaration of Human Rights*. www.un.org/en/about-us/universal-declaration-of-human-rights
- UN. (1966). *International Covenant on Civil and Political Rights*. General Assembly resolution 2200A (XXI). www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights
- UN. (2019). *United Nations Strategy and Plan of Action on Hate Speech*. United Nations Secretary-General. May. www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf
- UN. (2020). *Promoting Digital Public Goods*. United Nations Office of the Secretary General's Envoy on Technology.
- UN. (2022). *Countering Disinformation for the Promotion and Protection of Human Rights and Fundamental Freedoms Report of the Secretary-General*. United Nations General Assembly A/77/287. <https://digitallibrary.un.org/record/3987886?v=pdf>
- UN. (2023a). *Information Integrity on Digital Platforms*. Our Common Agenda Policy Brief 8. www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf
- UN. (2023b). *Interim Report: Governing AI for Humanity*. United Nations AI Advisory Board. www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf
- UN. (2023c). *UN Plan of Action on the Safety of Journalists and the Issue of Impunity*. CI-12/CONF.202.6.
- UN. (2024a). *Governing AI for Humanity: Final Report*. United Nations, AI Advisory Board. www.un.org/sites/un2.un.org/files/governing_ai_for_humaneity_final_report_en.pdf
- UN. (2024b). *Pact for the Future, Global Digital Compact, and Declaration on Future Generations*. www.un.org/sites/un2.un.org/files/sotf-pact_for_the_future_adopted.pdf
- UN. (2024c). *Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development*. General Assembly A/78/L.49. <https://digitallibrary.un.org/record/4043244?v=pdf>
- UN. (2024d). *United Nations Global Principles for Information Integrity: Recommendations for Multi-stakeholder Action*. www.un.org/sites/un2.un.org/files/un-global-principles-for-information-integrity-en.pdf
- UN. (2024e). What is hate speech? Understanding hate speech. www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech
- UNESCO. (2022a). *Global Standards for Media and Information Literacy, Curricula Development Guidelines*. www.unesco.org/sites/default/files/medias/files/2022/02/Global_Standards_for_Media_and_Information_Literacy_Curricula_Development_Guidelines_EN.pdf?hub=750
- UNESCO. (2022b). *K–12 AI Curricula: A Mapping of Government-Endorsed AI Curricula*. <https://unesdoc.unesco.org/ark:/48223/pf0000380602>
- UNESCO. (2022c). *Recommendation on the Ethics of Artificial Intelligence*. SHS/BIO/PI/2021/1. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- UNESCO. (2022d). *World Trends in Freedom of Expression and Media Development: Global Report 2021/22*.
- UNESCO. (2023a). *Addressing Hate Speech Through Education: A Guide for Policy Makers*. <https://unesdoc.unesco.org/ark:/48223/pf0000384872>
- UNESCO. (2023b). *Guidelines for the Governance of Digital Platforms: Safeguarding Freedom of Expression and Access to Information through a Multi-Stakeholder Approach*. <https://unesdoc.unesco.org/ark:/48223/pf0000387339>
- UNESCO. (2023c). *World Press Freedom Day 2023, Shaping a Future of Rights: Freedom of Expression as a Driver for All Other Human Rights; Draft Concept Note*. <https://unesdoc.unesco.org/ark:/48223/pf0000384177.locale=en>
- UNESCO. (2024). *Observatory of Killed Journalists*. www.unesco.org/en/safety-journalists/observatory
- Urman, A., & Makhortykh, M. (2024). The silence of the LLMs: Cross-lingual analysis of political bias and false information prevalence in ChatGPT, Google Bard, and Bing Chat. *OSFpreprints*, pp. 1–11. doi: 10.31219/osf.io/q9v8f.
- US Congress. (1996a). *Health Insurance Portability and Accountability Act (HIPAA)*. UPub. L. 104–191. <https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf>

- US Congress. (1996b). *Telecommunications Act of 1996, Section 230 Protection for Private Blocking and Screening of Offensive Material*. Pub. LA. 104-104. <https://www.govinfo.gov/app/details/PLAW-104publ104#:~:text=An%20act%20to%20promote%20competition,deployment%20of%20new%20telecommunications%20technologies>.
- US Congress. (2013). *Children's Online Privacy Protection Act 1998*. U15 USC 6501-6505, Federal Trade Commission Amendment. <https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa>
- US District Court. (2024). *US v. Google*. United States District Court for the District of Columbia, Case No. 20-cv-3010 (APM), 5 August. <https://www.texasattorneygeneral.gov/sites/default/files/images/press/Google%20Search%20Engine%20Monopoly%20Ruling.pdf>
- US Executive Order. (2023). *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. The White House. www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence
- US Government. (2020). *National Artificial Intelligence Initiative Act of 2020*. US Government H. R. 6395-1136, Division E. <https://www.aip.org/sites/default/files/aipcorp/images/fyi/pdf/national-ai-initiative-act-final.pdf>
- US State of California. (2018). *California Consumer Privacy Act (CCPA)*. State of California Department of Justice. <https://oag.ca.gov/privacy/ccpa>
- US-China Commission. (2022). *China's Evolving Data Governance Regime*. US-China Economic and Security Review Commission.
- Usher, N. (2023). The real problems with the problem of news deserts: Toward rooting place, precision, and positionality in scholarship on local news and democracy. *Political Communication*, 40(2), 238-253. <https://doi.org/10.1080/10584609.2023.2175399>
- Uyheng, J., & Carley, K. M. (2024). Online hate in the Philippines: The role of narratives and networks. *International Review of Law, Computers & Technology*, 38(2), 230-244. <https://doi.org/10.1080/13600869.2023.2295104>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1-13. <https://doi.org/10.1177/2056305120903408>
- Vaidhyanathan, S. (2011). *The Googlization of Everything: How One Company is Transforming Culture, Commerce, and Community – And Why We Should Worry*. University of California Press.
- Vaidhyanathan, S. (2022). *Anti-Social Media: How Facebook Disconnects Us and Undermines Democracy*. Oxford University Press.
- Valenzuela, S., Bachmann, I., Lawrence, R. G., & Gil De Zúñiga, H. (2023). Politics and media. *Journalism & Mass Communication Quarterly*, 100(4), 808-825. <https://doi.org/10.1177/10776990231203542>
- Valenzuela, S., Halpern, D., & Araneda, F. (2022). A downward spiral? A panel study of misinformation and media trust in Chile. *The International Journal of Press/Politics*, 27(2), 353-373. <https://doi.org/10.1177/19401612211025238>
- Valenzuela, S., Muñiz, C., & Santos, M. (2024). Social media and belief in misinformation in Mexico: A case of maximal panic, minimal effects? *The International Journal of Press/Politics*, 29(3), 646-666. <https://doi.org/10.1177/19401612221088988>
- van der Linden, S. (2023). *Foolproof: Why We Fall for Misinformation and How to Build Immunity*. Fourth Estate.
- Van der Spuy, A. (2023). *Unveiling Digital Development Risks: The Uncertain Outcomes of Promoting Digital Technologies in the Guise of Development*. PhD Thesis, London School of Economics and Political Science.
- van der Vlist, F., Helmond, A., & Ferrari, F. (2024). Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence. *Big Data & Society*, 11(1), 1-16. <https://doi.org/10.1177/20539517241232630>
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197-208. <https://doi.org/10.24908/ss.v12i2.4776>
- van Dijck, J. (2020). Seeing the forest for the trees: Visualizing platformization and its governance. *New Media & Society*, 23(9), 2801-2819. <https://doi.org/10.1177/1461444820940293>
- van Dijck, J., & Poell, T. (2013). Understanding social media logic. *Media and Communication*, 1(1), 2-14. <https://doi.org/10.12924/mac2013.01010002>
- van Dijck, J., de Waal, M., & Poell, T. (2018a). *The Platform Society: Public Values in a Connective World*. Oxford University Press.
- van Dijck, J., Poell, T., & de Waal, M. (2018b). News. In J. van Dijck, T. Poell, & M. de Waal, *The Platform Society: Public Values in a Connective World*. Oxford University Press, (pp. 49-72). Oxford University Press.
- van Drunen, M. Z., & Noroozian, A. (2024). How to design data access for researchers: A legal and software development perspective. *Computer Law & Security Review*, 52(2024), 1-15. <https://doi.org/10.1016/j.clsr.2024.105946>
- Van Natta, J., Masadeh, S., & Hamilton, B. (2023). Investigating the impacts of YouTube's content policies on journalism and political discourse. *Proceedings of the ACM on Human-Computer Interaction*, 7, CSCW1, 16 April. <https://doi.org/10.1145/3579499>
- Vanoli, C. (2024). Moving AI governance from principles to practice. *ITU News*, 19 April. www.itu.int/hub/2024/04/moving-ai-governance-from-principles-to-practice
- Vayadande, K., Singh, V., Sultanpure, K., Deshpande, D., et al. (2024). Empowering data sovereignty: Decentralized image sharing through blockchain and InterPlanetary File System. *International Journal of Intelligent Systems and Applications in Engineering*, 12(14S), 223-235. <https://ijisae.org/index.php/IJISAE/article/view/4660>

- Vázquez-Herrero, J., Negreira-Rey, M.-C., & López-García, X. (2022a). Let's dance the news! How the news media are adapting to the logic of TikTok. *Journalism*, 23(8), 1717–1735. <https://doi.org/10.1177/1464884920969092>
- Vázquez-Herrero, J., Negreira-Rey, M.-C., & Sixto-García, J. (2022b). Mind the gap! Journalism on social media and news consumption among young audiences. *International Journal of Communication*, 16(2022), 21. <https://ijoc.org/index.php/ijoc/article/view/19643>
- Veale, M., Matus, K., & Gorwa, R. (2023). AI and global governance: Modalities, rationales, tensions. *Annual Review of Law and Social Science*, 19, 255–275. <https://doi.org/10.1146/annurev-lawsocsci-020223-040749>
- Veiga, A. K., Saraiva, A. M., Chapman, A. D., Morris, P. J., et al. (2017). A conceptual framework for quality assessment and management of biodiversity data. *PLOS ONE*, 12(6), e0178731. <https://doi.org/10.1371/journal.pone.0178731>
- Veliz, C. (2021). *Privacy is Power: Why and How You Should Take Back Control of Your Data*. Melville House.
- Velkova, J. (2024). Dismantling public values, one data center at the time. NordMedia Network, 20 February. <https://nordmedianetwork.org/latest/news/dismantling-public-values-one-data-center-at-the-time>
- Ven, K., Verelst, J., & Mannaert, H. (2008). Should you adopt open source software? *IEEE Software*, 25(3), 54–59. <https://doi.org/10.1109/MS.2008.73>
- Verdegem, P. (2021). Tim Berners-Lee's plan to save the internet: Give us back control of our data. *The Conversation*, 5 February. <https://theconversation.com/tim-berners-lees-plan-to-save-the-internet-give-us-back-control-of-our-data-154130>
- Verdegem, P. (2022). Dismantling AI capitalism: The commons as an alternative to the power concentration of Big Tech. *AI & Society*, 39, 727–737. <https://doi.org/10.1007/s00146-022-01437-8>
- Verhulst, S. G. (2023). Steering responsible AI: A case for algorithmic pluralism. *ArXiv*. <https://doi.org/10.48550/arXiv.2311.12010>
- Verhulst, S. G. (2024). Governing Mediation in the Data Ecosystem: Lessons from Media Governance for Overcoming Data Asymmetries. In M. Puppis, R. Mansell, & H. Van den Bulck (Eds), *Handbook of Media and Communication Governance* (pp. 372–383). Edward Elgar Publishing.
- Verhulst, S. G., & Schüür, F. (2023). Interwoven realms: Data governance as the bedrock for AI governance. *Medium, Blog*, 20 November. <https://medium.com/data-policy/interwoven-realms-data-governance-as-the-bedrock-for-ai-governance-ffd56a6a4543>
- Verma, N., Fleischmann, K. R., & Koltai, K. S. (2018). Demographic factors and trust in different news sources. *Proceedings of the Association for Information Science and Technology*, 55(1), 524–533. <https://doi.org/10.1002/pra2.2018.14505501057>
- Verza, S., Blagojev, T., Da Cots Leite Borges, D., Kermer, J. E., Trevisan, M., & Reviglio Della Venaria, U. (2024). *Uncovering News Deserts in Europe: Risks and Opportunities for Local and Community Media in the EU*. European University Institute.
- Viljoen, S. (2021). A relational theory of data governance. *Yale Law Journal*, 131(2), 573–654. www.yalelawjournal.org/feature/a-relational-theory-of-data-governance
- Villi, M., Aharoni, T., Tenenboim-Weinblatt, K., Boczkowski, P. J., et al. (2022). Taking a break from news: A five-nation study of news avoidance in the digital era. *Digital Journalism*, 10(1), 148–164. <https://doi.org/10.1080/21670811.2021.1904266>
- Vincent, R. C., & Nordenstreng, K. (Eds) (2015). *Towards Equity in Global Communication? MacBride Update*. Second Edition. Hampton Press.
- Vinhas, O., & Bastos, M. (2022). Fact-checking misinformation: Eight notes on consensus reality. *Journalism Studies*, 23(4), 448–468. <https://doi.org/10.1080/1461670X.2022.2031259>
- Vogels, E. A., & Anderson, M. (2019). *Americans and Digital Knowledge*. Pew Research Center, US.
- von Ingersleben-Seip, N. (2023). Competition and cooperation in artificial intelligence standard setting: Explaining emergent patterns. *Review of Policy Research*, 40(5), 781–810. <https://doi.org/10.1111/ropr.12538>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Voss, W. G., & Pernot-Leplay, E. (2024). China data flows and power in the era of Chinese big tech. *Northwestern Journal of International Law and Business*, 44(1), 1–68. <https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1896&context=njilb>
- VPUU (Violence Prevention thorough Urban Upgrading). (2019). CitySpec: A mobile inspection tool improving service delivery. <http://vpuu.org.za/safe-node-area/cityspec-service-delivery>
- Vuorikari, R., Kluzer, S., & Punie, Y. (2022). *DigComp 2.2: The Digital Competence Framework for Citizens – With New Examples of Knowledge, Skills and Attitudes*. European Commission, Joint Research Centre.
- Wagner, B., Kettemann, M. C., Vieth-Dilmann, K., & Montgomery, S. (Eds) (2025 January). *Research Handbook on Human Rights and Digital Technology: Global Politics, Law and International Relations*. Second Edition. Edward Elgar Publishing.
- Wahl-Jorgensen, K., & Carlson, M. (2021). Conjecturing fearful futures: Journalistic discourses on deepfakes. *Journalism Practice*, 15(6), 803–820. <https://doi.org/10.1080/17512786.2021.1908838>
- Waisbord, S. (2023). Mob censorship revisited: Questions, findings, and challenges. *Digital Journalism*, 11(10), 1761–1768. <https://doi.org/10.1080/21670811.2023.2214905>
- Wakefield, J. (2016). Social media 'outstrips TV' as news source for young people. BBC News, 15 June. www.bbc.com/news/uk-36528256

- Waldman, A. E. (2020). Cognitive biases, dark patterns, and the 'privacy paradox'. *Current Opinion in Psychology*, 31, 105–109. <https://doi.org/10.1016/j.copsyc.2019.08.025>
- Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, 37(3), 350–375. <https://doi.org/10.1080/10584609.2019.1668894>
- WAN-IFRA (World Association of News Publishers). (2023). *Gauging Generative AI's Impact on Newsrooms: Survey: Newsroom Executives Share Their Experiences So Far*. World Association of News Publishers. <https://wan-ifra.org/insight/gauging-generative-ais-impact-in-newsrooms>
- Wang, B., Rau, P.-L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 42(9), 1324–1337. <https://doi.org/10.1080/0144929X.2022.2072768>
- Wang, C., Boerman, S. C., Kroon, A. C., Möller, J., & de Vreese, C. H. (2024). The artificial intelligence divide: Who is the most vulnerable? *New Media & Society, Online First*, 1–23. <https://doi.org/10.1177/14614448241232345>
- Wang, W. W. (2023). China's digital transformation: Data-empowered state capitalism and social governmentality. *The African Journal of Information and Communication*, 2023(31), 1–13. <https://doi.org/10.23962/ajic.i31.16296>
- Wardle, C. (2018). The need for smarter definitions and practical, timely empirical research on information disorder. *Digital Journalism*, 6(8), 951–963. <https://doi.org/10.1080/21670811.2018.1502047>
- Wardle, C. (2023). Misunderstanding misinformation. *Issues in Science and Technology*, 29(3), 38–40. <https://doi.org/10.58875/ZAUD1691>
- Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Council of Europe Report DGO(2017)09. <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>
- Warschauer, M. (2004). *Technology and Social Inclusion: Rethinking the Digital Divide*. MIT Press.
- Washington, J. (2023). *Combating Misinformation and Fake News: The Potential of AI and Media Literacy Education*. <http://dx.doi.org/10.2139/ssrn.4580385>
- Wasko, J., Murdock, G., & Sousa, H. (Eds) (2011). *The Handbook of Political Economy of Communications*. John Wiley & Sons, Ltd.
- Wasserman, H. (2018). *Media, Geopolitics, and Power: A View from the Global South*. University of Illinois Press.
- Wasserman, H. (2020a). Fake news from Africa: Panics, politics and paradigms. *Journalism*, 21(1), 3–16. <https://doi.org/10.1177/1464884917746861>
- Wasserman, H. (2020b). The state of South African media: A space to contest democracy. *Publizistik*, 65(3), 451–465. <https://doi.org/10.1007/s11616-020-00594-4>
- Wasserman, H. (Ed.). (2022). *Meeting the Challenges of Information Disorder in the Global South*. IDRC (International Development Research Centre, Canada) Research Results.
- Wasserman, H., & Madrid-Morales, D. (2019). An exploratory study of 'fake news' and media trust in Kenya, Nigeria and South Africa. *African Journalism Studies*, 40(1), 107–123. <https://doi.org/10.1080/23743670.2019.1627230>
- WEF (World Economic Forum). (2024). *Global Risks Report 2024*. 19th Edition. Insight Report. www.weforum.org/publications/global-risks-report-2024
- WEF (World Economic Forum). (2024). How to manage AI's energy demand – today and in the future. 25 April. www.weforum.org/agenda/2024/04/how-to-manage-ais-energy-demand-today-tomorrow-and-in-the-future
- Weismueller, J., Gruner, R. L., Harrigan, P., Coussement, K., & Wang, S. (2023). Information sharing and political polarisation on social media: The role of falsehood and partisanship. *Information Systems Journal, Special Issue* (2023), 1–40. <https://doi.org/10.1111/isj.12453>
- Werthner, H., Ghezzi, C., Kramer, J., Nida-Rümelin, J., et al. (Eds) (2024). *Introduction to Digital Humanism: A Textbook*. Springer Nature Switzerland.
- West, E., & McAllister, M. P. (Eds) (2023). *The Routledge Companion to Advertising and Promotional Culture*. Taylor & Francis.
- West, S. M. (2019). Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society*, 58(1), 20–41. <https://doi.org/10.1177/0007650317718185>
- Whittaker, M. (2021). The steep cost of capture. *Interactions*, 28(6), 50–55. <https://doi.org/10.1145/3488666>
- Wiggers, K. (2019). IBM releases Diversity in Faces, a dataset of over 1 million annotations to help reduce facial recognition bias. VentureBeat, 29 January. <https://venturebeat.com/ai/ibm-releases-diversity-in-faces-a-dataset-of-over-1-million-annotations-to-help-reduce-facial-recognition-bias>
- Wihbey, J. (2024). AI and epistemic risk for democracy: A coming crisis of public knowledge? <https://doi.org/10.2139/ssrn.4805026>
- Wike, R., Fetterolf, J., Smerkovich, M., Austin, S., Gubbala, S., & Lippert, J. (2024). Representative democracy remains a popular ideal, but people around the world are critical of how it's working. Pew Research Center Press Release, 28 February. www.pewresearch.org/global/wp-content/uploads/sites/2/2024/02/gap_2024.02.28_democracy-closed-end_report.pdf
- Wikipedia. (2024). Sock puppet account. https://en.wikipedia.org/wiki/Sock_puppet_account

- Wilding, D. (2021). Regulating news and disinformation on digital platforms: Self-regulation or prevarication? *Journal of Telecommunications and the Digital Economy*, 9(2), 11–46. <https://doi.org/10.18080/jtde.v9n2.415>
- Willems, W. (2014a). Beyond normative dewesternization: Examining media culture from the vantage point of the Global South. *The Global South*, 8(1), 7–23. <https://doi.org/10.2979/globalsouth.8.1.7>
- Willems, W. (2014b). Provincializing hegemonic histories of media and communication studies: Toward a genealogy of epistemic resistance in Africa. *Communication Theory*, 24(4), 415–434. <https://onlinelibrary.wiley.com/doi/abs/10.1111/comt.12043>
- Williams, A. (2023). *Surveillance Nation: The Real Cost of Amazon's Customer Obsession*. <https://peertube.dair-institute.org/w/oAWnan9PXTvS3CyutatQcg>
- Windwehr, S., & York, J. C. (2020). Facebook's most recent transparency report demonstrates the pitfalls of automated content moderation. Electronic Frontier Foundation, 8 October. www.eff.org/deeplinks/2020/10/facebooks-most-recent-transparency-report-demonstrates-pitfalls-automated-content
- Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record*, 121(11), 1–40. <https://doi.org/10.1177/016146811912101102>
- Winseck, D. (2022). The Broken Internet and Platform Regulation: Promises and Perils. In T. Flew & F. R. Martin (Eds), *Digital Platform Regulation: Global Perspectives on Internet Governance* (pp. 229–257). Springer International Publishing.
- Winseck, D., & Pooley, J. D. (2017). A reply to Faulhaber, Singer, and Urschel's Curious tale of economics and common carriage (net neutrality) at the FCC. *International Journal of Communication*, 11(0), 32. <https://ijoc.org/index.php/ijoc/article/view/7543>
- Wintour, P. (2024). US leading global alliance to counter foreign government disinformation. *The Guardian*, 26 February. www.theguardian.com/technology/2024/feb/26/us-leading-global-alliance-to-counter-foreign-government-disinformation
- WIPO (World Intellectual Property Organization). (2024). *Getting the Innovation Ecosystem Ready for AI: An IP Policy Toolkit*. Frontier Technologies Division. www.wipo.int/publications/en/details.jsp?id=4711
- Wojcowski, T. (2022). Conspiracy theories, right-wing populism and foreign policy: The case of the Alternative for Germany. *Journal of International Relations and Development*, 25(2022), 130–148. <https://doi.org/10.1057/s41268-021-00218-y>
- Wood, S. (2024). *Impact of Regulation on Children's Digital Lives*. Digital Futures for Children, London School of Economics and Political Science and 5Rights Foundation.
- Wu, C.-J., Raghavendra, R., Gupta, U., Acun, B., et al. (2022). Sustainable AI: Environmental implications, challenges and opportunities. ArXiv. <https://doi.org/10.48550/arXiv.2111.00364>
- Wu, J., Gan, W., Chen, Z., Wan, S., & Lin, H. (2023). AI-generated content (AIGC): A survey. ArXiv. <https://arxiv.org/pdf/2304.06632>
- Wu, T. (2003). Network neutrality, broadband discrimination. *Journal of Telecommunications and High Technology Law*, 2, 141–179. <https://doi.org/10.2139/ssrn.388863>
- Wu, T. (2017). *The Attention Merchants: The Epic Struggle to Get Inside Our Heads*. Atlantic Books.
- Wu, T. (2018). *The Curse of Bigness: Antitrust in the New Gilded Age*. Columbia Global Reports.
- Wulf, A. I. J., & Seizov, O. (2022). 'Please understand we cannot provide further information': Evaluating content and transparency of GDPR-mandated AI disclosures. *AI & SOCIETY*, 39, 235–256. <https://doi.org/10.1007/s00146-022-01424-z>
- Wuyckens, G., Landry, N., & Fastrez, P. (2022). Untangling media literacy, information literacy, and digital literacy: A systematic meta-review of core concepts in media education. *Journal of Media Literacy Education*, 14(1), 168–182. <https://doi.org/10.23860/JMLE-2022-14-1-12>
- X. (2024). About Community Notes on X. X Help Center. <https://help.x.com/en/using-x/community-notes>
- Xu, J. (2023). What does China's newly launched National Data Bureau mean to China and global data governance? *Internet Policy Review*, 25 April. <https://policyreview.info/articles/news/chinas-national-data-bureau-and-global-data-governance>
- Xu, Z., Jain, S., & Kankanhalli, M. (2024). Hallucination is inevitable: An innate limitation of large language models. ArXiv. <https://doi.org/10.48550/arXiv.2401.11817>
- Yan, B., Li, K., Xu, M., Dong, Y., et al. (2024). On protecting the data privacy of large language models (LLMs): A survey. ArXiv. <https://doi.org/10.48550/arXiv.2403.05156>
- Yang, Y., & Bennett, L. (2022). Interactive Propaganda: How Fox News and Donald Trump Co-Produced False Narratives about the COVID-19 Crisis. In P. Van Aelst & J. G. Blumler (Eds), *Political Communication in the Time of Coronavirus* (pp. 83–100). Routledge.
- Yates, S., & Carmi, E. (Eds) (2024). *Digital Inclusion: International Policy and Research*. Springer International Publishing.
- Ye, Q. (2023). Comparison of the transparency of fact-checking: A global perspective. *Journalism Practice*, 17(10), 2263–2282. <https://doi.org/10.1080/17512786.2023.2211555>
- Yoo, C. S. (2024). Network slicing and net neutrality. *Telecommunications Policy*, 48(2), 1–10. <https://doi.org/10.1016/j.telpol.2023.102619>
- York, J. C. (29 October 2022). Elon Musk doesn't know what it takes to make a digital town square. MIT Technology Review, 29 October. www.technologyreview.com/2022/10/29/1062417/elon-musk-twitter-takeover-global-democracy-activists

- Ytre-Arne, B., & Moe, H. (2021). Doomscrolling, monitoring and avoiding: News use in COVID-19 pandemic lockdown. *Journalism Studies*, 22, 1739–1775. www.tandfonline.com/doi/full/10.1080/1461670X.2021.1952475
- Zammit, M., Voulgari, I., & Yannakakis, G. N. (2021). The Road to AI Literacy Education: From Pedagogical Needs to Tangible Game Design. University of Malta Paper for European Conference on Games Based Learning. www.academia.edu/78830274/The_Road_to_Ai_Literacy_Education_from_Pedagogical_Needs_to_Tangible_Game_Design
- Zelizer, B. (2019). Why journalism is about more than digital technology. *Digital Journalism*, 7(3), 343–350. <https://doi.org/10.1080/21670811.2019.1571932>
- Zhang, W., Ong, L. M., & Findlay, M. (2023). Digital Self-Determination: An Alternative Paradigm for Emerging Economies. In M. Findlay, L. M. Ong, & W. Zhang (Eds), *Elgar Companion to Regulating AI and Big Data in Emerging Economies* (pp. 158–179). Edward Elgar Publishing.
- Zhang, Y., Lukito, J., Su, M.-H., Suk, J., et al. (2021). Assembling the networks and audiences of disinformation: How successful Russian IRA Twitter accounts built their followings, 2015–2017. *Journal of Communication*, 71(2), 305–331. <https://doi.org/10.1093/joc/jqaa042>
- Zhao, Y., & Chen, J. (2022). A survey on differential privacy for unstructured data content. *ACM Computing Surveys*, 54(10s), 207, 1–28. <https://doi.org/10.1145/3490237>
- Zhu, J. (2022). AI ethics with Chinese characteristics? Concerns and preferred solutions in Chinese academia. *AI & Society*, 39(3), 1261–1274. <https://doi.org/10.1007/s00146-022-01578-w>
- Zhu, Q., Weeks, B. E., & Kwak, N. (2024). Implications of online incidental and selective exposure for political emotions: Affective polarization during elections. *New Media & Society*, 26(1), 450–472. <https://doi.org/10.1177/14614448211061336>
- Zibi Fama, P. A. (2024). *La démocratie à l'épreuve de la désinformation genrée en Afrique centrale et de l'Ouest*. Organisation Internationale de la francophonie (ODIL).
- Znojek, B. (2020). The growing problem of disinformation in Latin America. *PISM Bulletin*. PISM (Polski Instytut Spraw Międzynarodowych).
- Zuazo, N., & Aruguete, N. (2021). ¿Polarización política o digital? un ecosistema con todos los climas. In L. A. Quevedo & I. Ramirez (Eds), *Polarizados: ¿Por qué preferimos la grieta? (aunque digamos lo contrario)* (pp. 135–154). Capital Intelectual.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.
- Zullo, C. (2023). Addressing gender-based violence in the Central Sahel. Burkina Faso Reliefweb. InterAction, 5 December. www.interaction.org/blog/addressing-gender-based-violence-in-the-central-sahel
- Zygmuntowski, J. J., Zoboli, L., & Nemitz, P. F. (2021). Embedding European values in data governance: A case for public data commons. *Internet Policy Review*, 10(3), 1–29. <https://doi.org/10.14763/2021.3.1572>

Supported by

