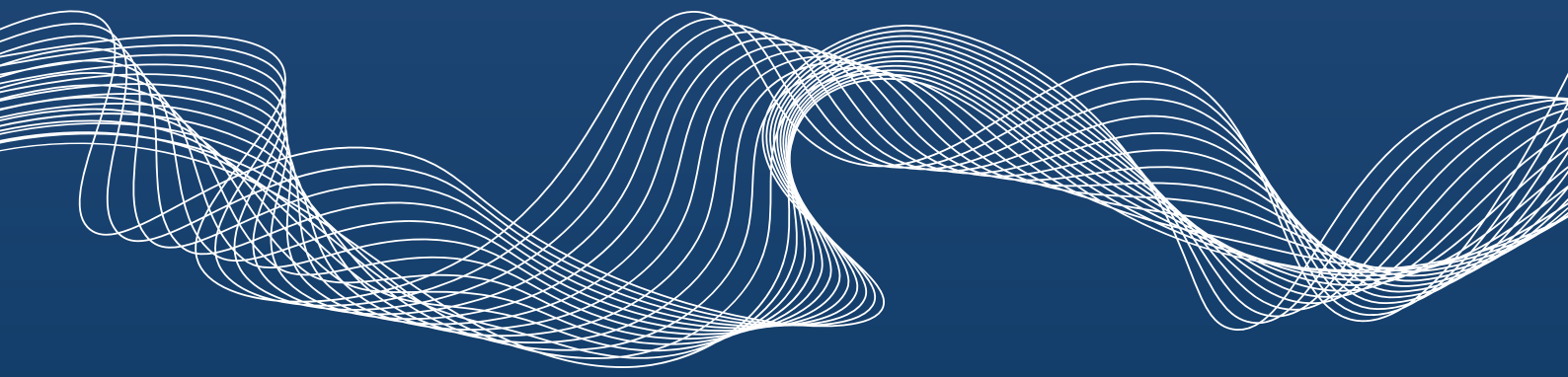
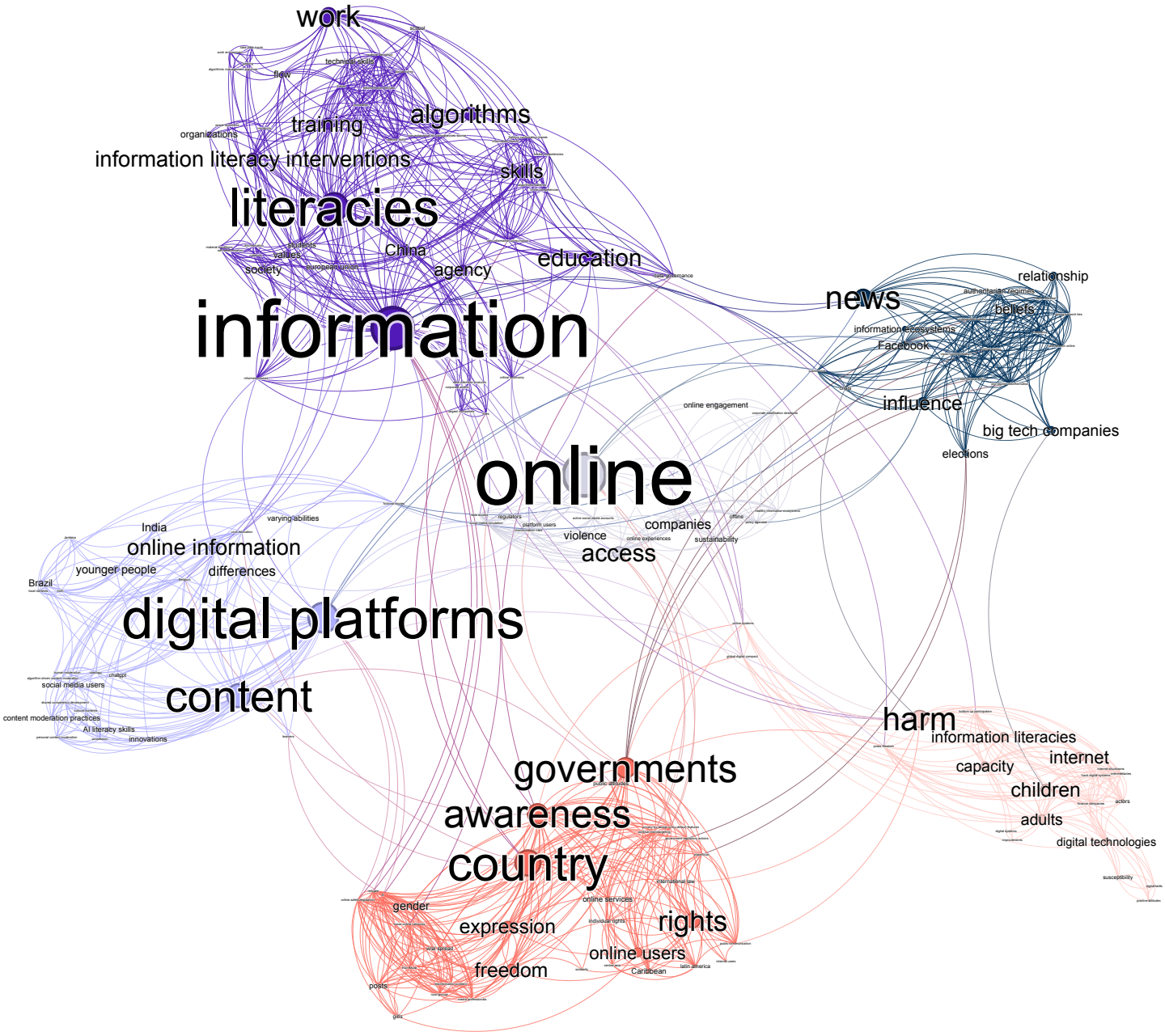
OBSERVATORY ON INFORMATION AND DEMOCRACY

AWARENESS OF MIS- AND DISINFORMATION AND THE LITERACY CHALLENGE



An initiative by: FORUM ON INFORMATION AND DEMOCRACY

How to cite this document: Mansell, R., Durach, F., Kettemann, M., Lenoir, T., Procter, R., Tripathi, G., and Tucker, E. (2025) 'Chapter 5: Awareness of Mis- and Disinformation and the Literacy Challenge' in 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.



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' (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).
[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:

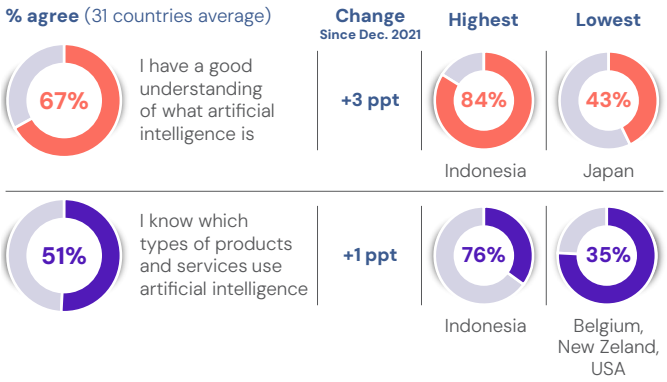
> 1) the majority of regular internet users globally (58.5%) worry about misinformation, and young and low-income groups are most likely to be concerned. 2) Risk perception among internet users *varies starkly across regions* whereby concern is highest in Latin America and the Caribbean (74.2%), and 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).

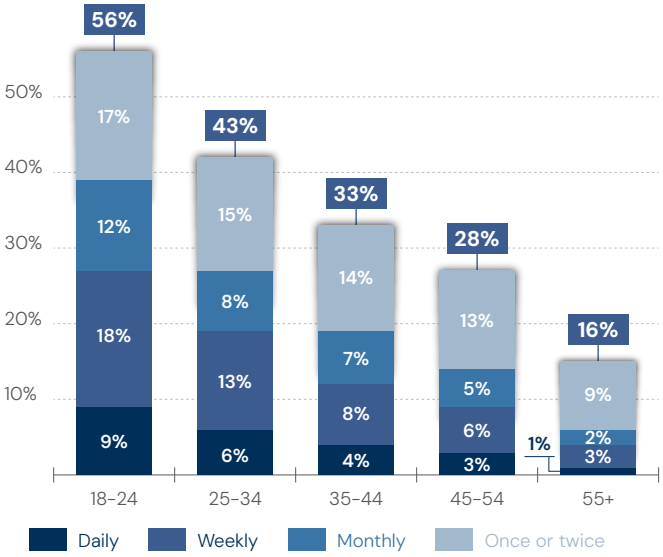
[33] Kyriakidou *et al.* (2023), supported by the Arts and Humanities Research Council (AHRC), UK.

[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 makes 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 online 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]

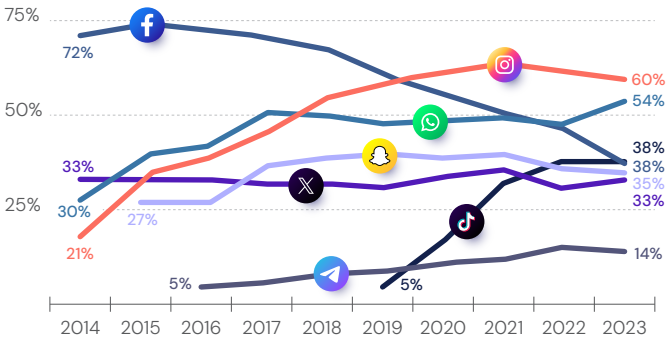
- In the *United States* predictions of how people respond to algorithm-driven content moderation and selection were found to depend on the

[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]

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]
- 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.

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]

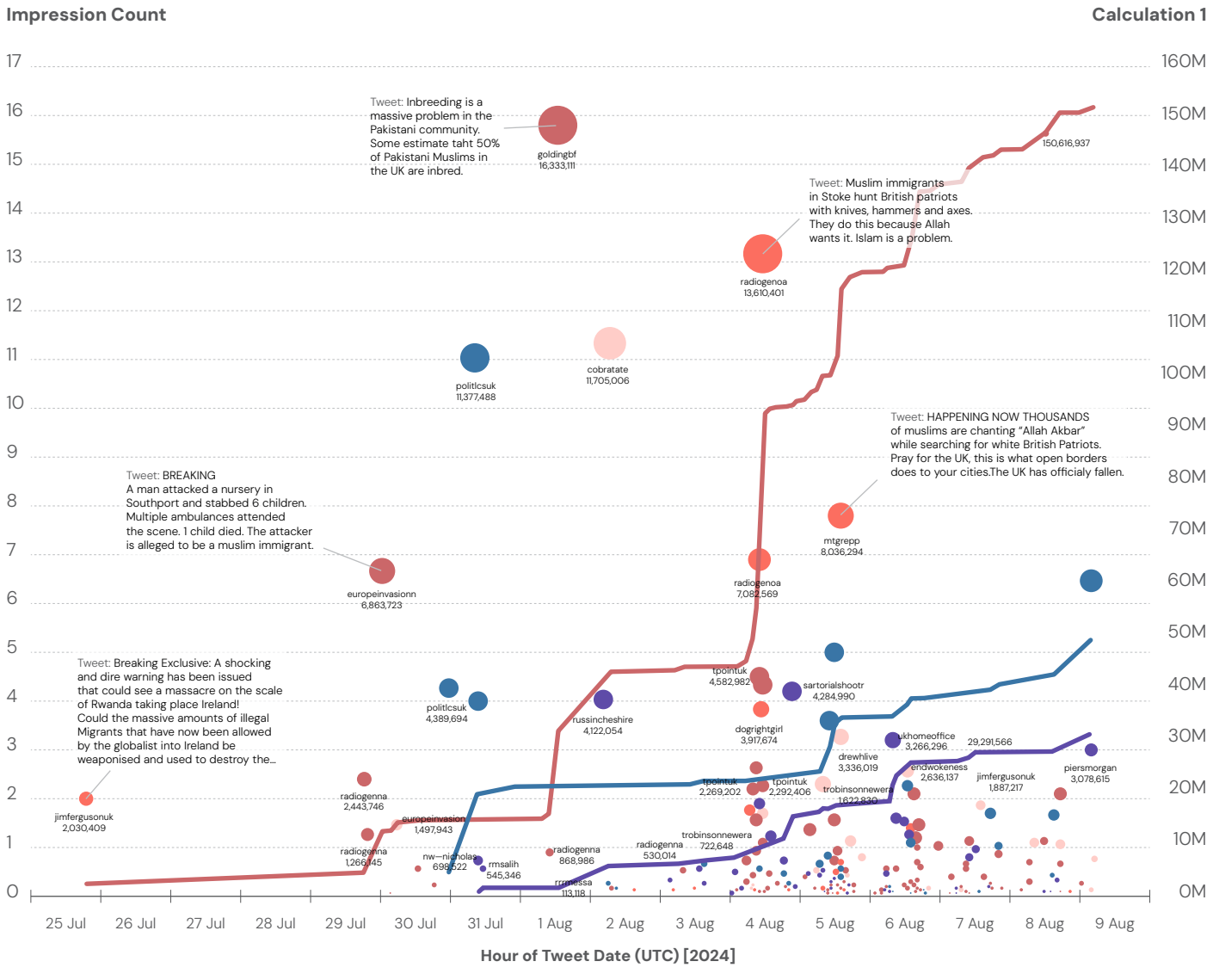
Viral spread of false information. In the United Kingdom, false claims posted on X that the killer of three young girls in the English town of Southport in July 2024 was a 17-year-old asylum seeker who had arrived in the country on a boat in 2023 were quickly followed by a wave of riots involving far-right 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 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 online 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.

References

- 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
- ADRN (African Digital Rights Network). (2024). *The Digitalisation of Social Protection in Africa: The Perspective of Marginalied 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
- 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
- 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/10.1332/030557321X16873376680333
- Banaji, S., & Bhat, R. (2022). *Social Media and Hate*. Routledge.
- 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.
- 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.
- 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
- 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
- Bradshaw, S., & DeNardis, L. (2022). Internet Infrastructure as an Emerging Terrain of Disinformation. CIGI (Centre for International Governance Innovation), 4 July. www.cigionline.org/articles/internet-infrastructure-as-an-emerging-terrain-of-disinformation
- 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
- Bunting, M. (2018). *Keeping Consumers Safe Online: Legislating for Platform Accountability for Online Content*. Communications Chambers.
- 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.
- 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
- 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
- Casciani, D., & BBC Verify. (2024). Violent Southport protests reveal far-right organising tactics. BBC News, 2 August. www.bbc.com/news/articles/cl4y0453nv5o
- 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
- 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
- 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
- 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
- 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.

- 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
- 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/e019014
- 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
- De Abreu, B. S. (Ed.) (2022). *Media Literacy, Equity, and Justice*. Taylor & Francis.
- 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
- 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
- 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.
- Dugan, L., Ippolito, D., Kirubarajan, 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
- Dunn, H. S., Ragnedda, M., Ruiu, M. L., & Robinson, L. (Eds) (2024). *The Palgrave Handbook of Everyday Digital Life*. Palgrave Macmillan.
- 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.
- EC (European Commission). (2017). *Mapping of Media Literacy Practices and Actions in EU-28*. European Commission Audio-Visual Observatory.
- 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
- Elsom, C. (2020). *Safety without Censorship: A Better Way to Tackle Online Harms*. Centre for Policy Studies, London.
- 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.
- Faculty. (2021). *Automated Approaches to Measuring Online Experiences: Executive Summary*. www.bleadon.org.uk/media/other/24400/OFCOM-automated-tooling-report-facultyreport.pdf
- 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., Neundorf, 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
- 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.
- 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
- 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. (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.
- Freire, P. (1974). Education for Critical Consciousness. Continuum.
- Friesem, Y., Raman, U., Kanižaj, I., & Choi, G. Y. (Eds) (2022). *The Routledge Handbook of Media Education Futures Post-Pandemic*. Routledge.
- 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).
- 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
- 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
- Grizzle, A., Wilson, C., Tuazon, R., Cheung, C. K., *et al.* (2021). *Media and Information Literate Citizens: Think Critically*, Click Wisely! UNESCO.
- Haider, J., & Sundin, O. (2022). *Paradoxes of Media and Information Literacy: The Crisis of Information*. Taylor & Francis.
- 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
- 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
- Howard, P. N., Neudert, L.-M., Prakash, N., & Vosloo, S. (2021). *Digital Misinformation / Disinformation and Children*. UNICEF Rapid Analysis Report.
- 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
- 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
- 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
- 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*.
- 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
- 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
- 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 (International Development Research Centre, Canada) 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
- 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
- Jones, M. O. (2024). News analysis all tweets mentioning both 'Muslim' & 'UK'. X (formerly Twitter), 9 August. https://x.com/marcowenjones/status/1821881958537871568
- 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
- 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, 612016), 27–35. https://doi.org/10.1016/j.chb.2016.02.080
- 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
- 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.
- 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
- 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
- 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
- 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

- 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
- 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. (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., & Blum-Ross, A. (2020). *Parenting for a Digital Future: How Hopes and Fears about Technology Shape Children's Lives*. Oxford University Press.
- 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.
- 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
- 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
- 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
- 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.
- 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
- 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., 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
- 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.
- 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. (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
- 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
- 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
- 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, 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
- 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
- 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
- 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/pra2.487
- 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

- Nyamnjoh, F. B. (2006). *Insiders and Outsiders: Citizenship and Xenophobia in Contemporary Southern Africa*. Zed Books.
- 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. (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
- 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.
- 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
- 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
- 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
- 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
- Ragnedda, M., & Gladkova, A. (Eds) (2020). *Digital Inequalities in the Global South*. Springer Nature.
- 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
- RIA (Research ICT Africa). (2023b). *Information Integrity on Digital Platforms*. RIA input on United Nations Policy Brief 8, led by G. Berger.
- 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
- 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.
- 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
- 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., 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.
- 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
- Sádaba, C., Salaverría, R., & Bringué, 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
- Schmarzo, B. (2023). *AI & Data Literacy: Empowering Citizens of Data Science*. Packt Publishing Ltd.
- 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
- 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

- 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.
- 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
- 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
- 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
- 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
- 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
- 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 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
- 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
- UN (United Nations). (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. (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
- 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
- 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. (2022d). *World Trends in Freedom of Expression and Media Development: Global Report 2021/22*.
- 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
- 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 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.
- 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 Reasearch Centre.
- 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
- Wasserman, H. (Ed.). (2022). *Meeting the Challenges of Information Disorder in the Global South*. IDRC (International Development Research Centre, Canada) Research Results.
- 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
- 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.

- 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
- 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).
- 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