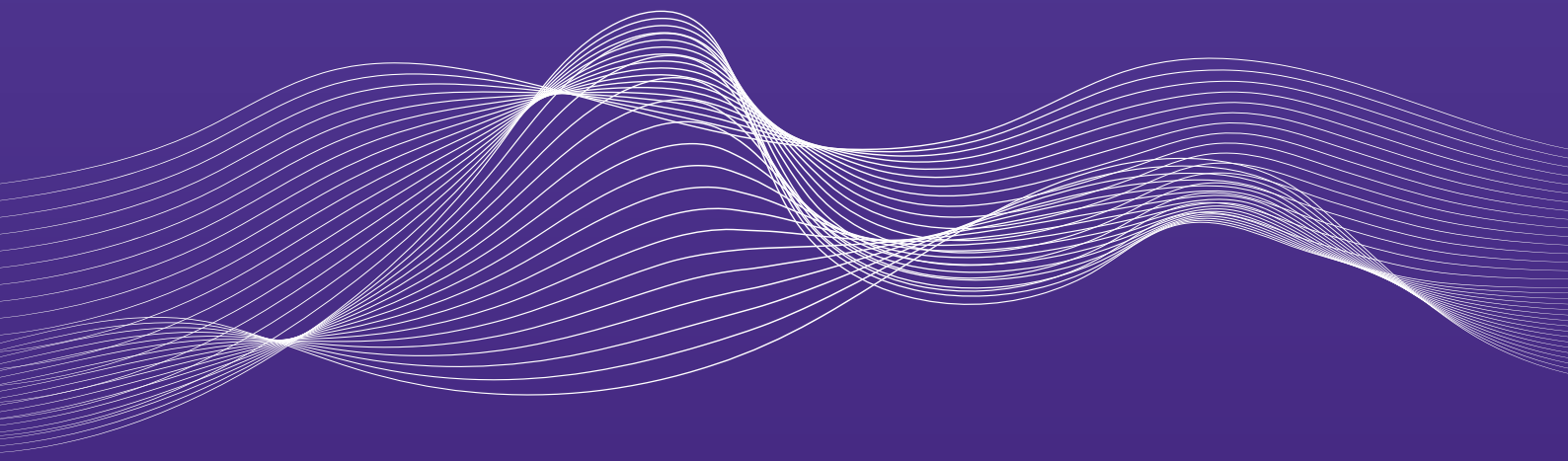
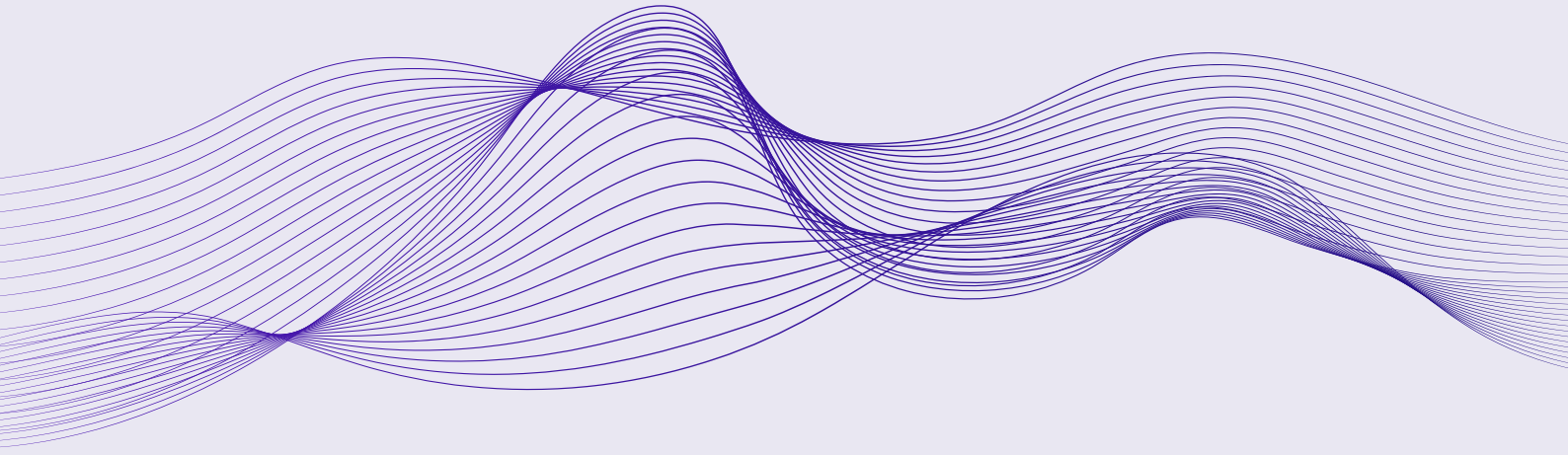
OBSERVATORY ON INFORMATION AND DEMOCRACY

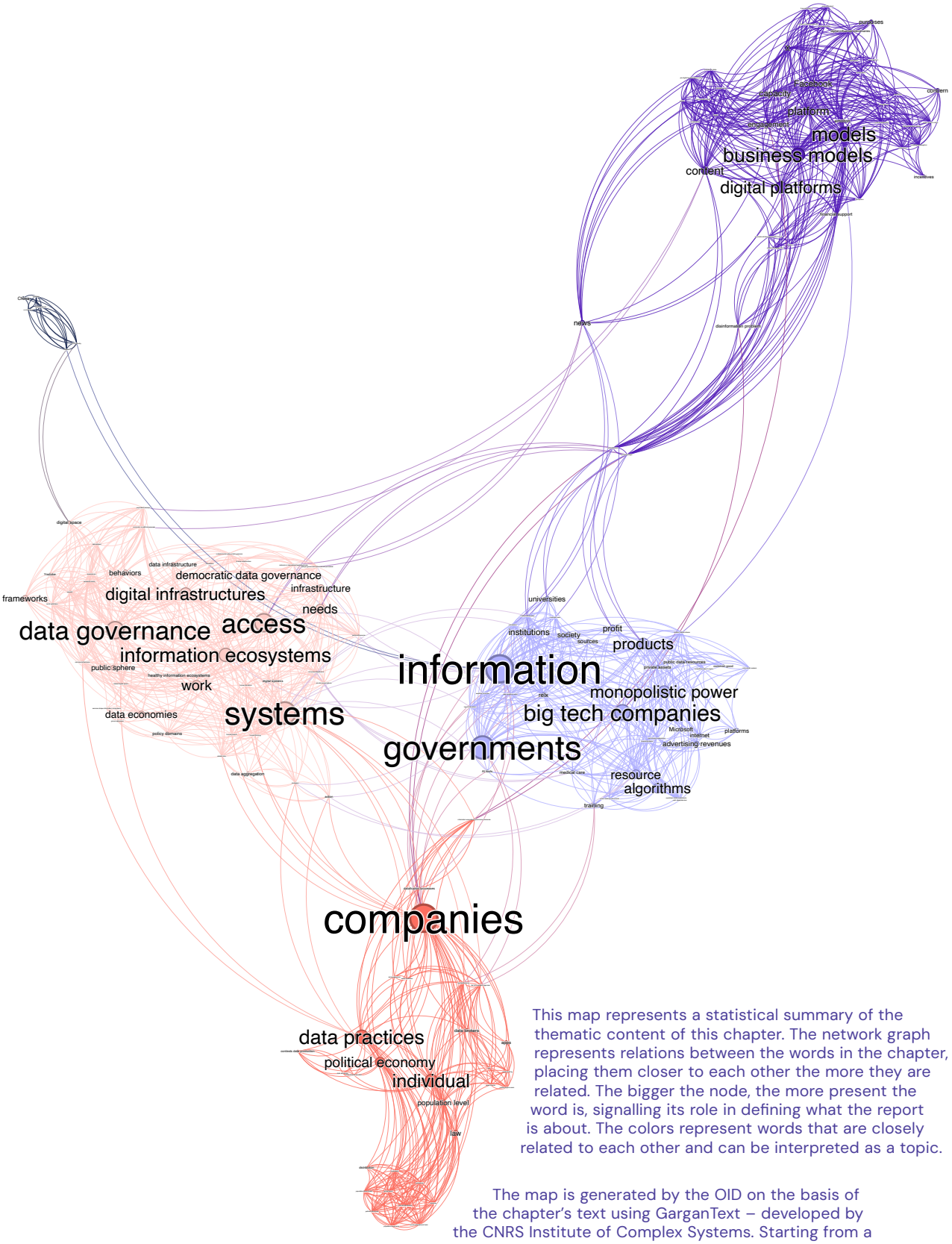
BIG TECH POWER AND GOVERNING USES OF DATA



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 4: Big Tech Power and Governing Uses of Data' 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 examines evidence on the relationships between the power of big tech companies and approaches to governing the practices of data extraction and use – that is, processes of datafication.[1] It examines approaches to governing uses of data that influence data practices - the generation and the uses to which data is put - which allow information ecosystems to exist and how these practices are experienced by individuals, communities and across all sectors of economies.

The research synthesis focuses on:

- **What is the appropriate role of data and digital infrastructures within political communities?** This examines research on how to govern the massive amounts of data that are the raw material of the digital economy. Why the design and operation of data and digital infrastructures are contested is examined in the light of big tech company practices to show why these practices are inconsistent with democracy.
- **How are data aggregation and AI systems changing the way people build, share and receive information and knowledge?** This focuses on the power of big tech companies to exert monopolistic control over data through their data extraction models. It explains how this leads to data being used in ways that create dependencies of individuals, communities and industry sectors on datafication processes. It discusses how big tech business strategies result in control over information that restricts access especially for people in the Global Majority World. It highlights injustices associated with the interplay of data mining and data brokering, explaining why digital platforms are incentivized to turn a 'blind eye' to mis- and disinformation.
- **How do these big tech strategies and practices interfere with political deliberation which is essential for the survival of participatory democracy?** This explains why it is not sufficient to examine the harms of datafication in individualistic terms. The focus is on how data practices produce or entrench social injustices at the population level including wealth disparities and racial oppression. The need for alternative approaches to democratic data governance is discussed with a critique of measures that leave the business models and data practices of big tech companies largely in place.

This chapter provides an assessment of research in these areas providing insight into the political economy of datafication processes.

In the next chapter (Chapter 5), the scale of the mis- and disinformation problem and what the public and policy makers understand about algorithmic-driven datafication systems is discussed. Chapter 5 also examines initiatives to strengthen individuals' capacities to control their own engagement with data-driven systems through media and information literacy as well as AI literacy. Chapters 6 and 7 discuss information ecosystem governance measures applied by governments and companies. Further discussion of data extractive practices is in Chapter 8 which critically examines alternative data governance practices.

[1] For background reading, see Aaronson (2021); Aguerre *et al.* (2024); Padovani *et al.* (2024); Taylor *et al.* (2022); Verhulst & Schüür (2023). See Appendix: Methodology for details of literature review process.

1 Introduction

The 'information ecosystem' is a metaphor. It is useful imagery for thinking about the dynamic, complex and interconnected set of systems that determines who can receive or share what kinds of information and in what contexts. Extending the metaphor, information ecosystems are shaped fundamentally by (and shape) the environments in which they arise – in today's context, a data-rich environment.[2] An understanding of how and why mis- and disinformation arise and spread requires us to understand the political economy of data – that is, the way power relations establish the conditions for how data economies operate, and how they are inequitably experienced by different people around the world.

Data is not a naturally occurring resource, however, and analogies with natural ecosystems distort our understanding of how data aggregation and algorithmic technologies alter the systems through which people build, share and receive information. Data is produced as a result of decisions by one or more human actors to create a record of something, such as an action performed by a human, usually for a particular purpose.[3] Data that constitutes today's information ecosystems is produced and controlled primarily by a small number of companies in accordance with business models designed to prioritize profit over corporate responsibility for human rights, privacy and safety.[4] Lurking behind the proliferation of mis- and disinformation (and other kinds of low-quality information), and the inaccessibility of useful and high-quality information, is the problem of *data governance practices* that are not designed to serve a democratically-arrived-at vision of how data should, or should not, shape the public sphere. This chapter explains why information integrity is at risk under current data governance arrangements, and why present developments in AI systems work against the requirements for healthy information ecosystems.

The OECD defines 'data governance' as the:

> Diverse arrangements, including technical, policy, regulatory and institutional provisions, that affect data and their creation, collection, storage, use, protection, access, sharing and deletion, including across policy domains and organisational and national borders. Efforts to govern data take many forms. They often seek to maximise the benefits from data, while addressing related risks and challenges, including to rights and interests.[5]

This definition mentions both benefits and risks. It positions data governance as being as much concerned with ensuring that data is used to drive economic growth and to favor corporate interests in data monetization as it is with protecting fundamental rights. It is, however, agnostic about whether the monopolistic activities of big tech companies, including corporate practices for data capture and control, are consistent with the protection of rights. In this chapter we investigate the injustices accompanying big tech company monopolistic behaviors and consider what data governance measures are needed to ensure the uses of data generated by digital systems and applications become more closely aligned with fairness and equality.

2 Digital Infrastructure Contestations

In the 1950s Hannah Arendt worried about a near future in which human technology would replace human thinking. In *The Human Condition*, she worried not because machines would become 'intelligent', but because of the many things machines would make it possible for human beings to do without 'intelligence':

[2] This perspective is consistent with a socio-technical view of the interpenetration of technology and society. See Chapter 1.

[3] Rosenberg (2013).

[4] De-Lima Santos (2023), funded by the University of Amsterdam and European Union Horizon 2020. See Section 2, Chapter 2 for a discussion of digital platform company incentives in relation to the news media component of information ecosystems.

[5] OECD (2022a, p. 13).

> [I]t could be that we, who are earth-bound creatures and have begun to act as though we are dwellers of the universe, will forever be unable to understand, that is, to think and speak about the things which nevertheless we are able to do. In this case, it would be as though our brain, which constitutes the physical, material condition of our thoughts, were unable to follow what we do, so that from now on we would indeed need *artificial machines* to do our thinking and speaking. If it should turn out to be true that knowledge (in the sense of know-how) and thought have parted company for good, then we would indeed become the helpless slaves, not so much of our machines as of our know-how, thoughtless creatures at the mercy of every gadget which is technically possible, no matter how murderous it is.[6]

One way Arendt's fear has been realized is by creating infrastructures for our social, political, cultural and economic systems that are pervasive, determinative and invisible.[7] Digital infrastructure is not disembodied; it involves familiar forms of industrial infrastructure including 'data centres distributed throughout the world and made up of servers, routers, switches, and miles of cables, as well as redundant power sources, cooling and ventilation systems, and security apparatus'.[8] Digital infrastructure is material and just as transfiguring of physical landscapes as railroads, highways and power grids. The difference is that the digital systems that 'shape, enable and sometimes deliberately constrain life in common'[9] are largely hidden from the conscious experience of people who depend on digital infrastructure for every aspect of life. The more that we depend on this infrastructure – to get from place to place, to shop, to access government services, to work or go to school, to get medical care and to have private conversations with friends – the more we become part of it. The more we become part of this infrastructure, the less we are aware of it, and the more it then shapes our perceptions of everything.[10]

The risks of engaging with the digital environment as if it is a natural environment vary depending on the context (there may be few downsides, such as obeying traffic lights within a well-designed algorithmic road safety system).[11] However, the consequences of the use of digital structures and data are hugely politically significant in our systems for creating, sharing and disseminating information.[12] Much academic research and civil society advocacy examining the dysfunction of today's information ecosystems focuses on algorithmic systems. However, to think creatively about what it will take to build healthy information ecosystems, it is essential to examine the fundamental problem of how to govern the creation of, access to, and use of massive amounts of data that is the raw material of all the digital economy. Our focus here is on research that offers critical perspectives on the forces resulting in the way data is used in today's digital information ecosystems and the prospects for supporting – or even enhancing – democratic governance.

These prospects require it to be feasible for polities to contest the design of systems and the mechanisms for controlling the users of data that makes these systems possible. Digital data-dependent systems already define and constrain political discourse and activity in many contexts. For those regions of the world that have not had the opportunity for substantial input into developing today's digital economy, it is especially crucial to acknowledge parallel and often conflicting visions for the governance of both the generation and uses of data.[13] It is also essential to recognize intersectional perspectives, including gender-sensitive approaches to data governance, and how they couple with other dimensions of

[6] Arendt ([1958] 1998; emphasis added).
[7] Cohen (2023); Star & Ruhleder (1996), part-funded by NSF US.
[8] Sacasas (2021).
[9] Stiefel *et al.* (2024).
[10] Barba-Kay (2023); McLuhan (1964); Morozov (2013); Zuboff (2019).
[11] There may be a variety of risks associated with algorithmic traffic management systems, such as privacy infringement, especially where system maintenance depends on multilayer real-time surveillance. See, for example, Local Progress (2024).
[12] Herman & Chomsky (1989). The English word data is the plural of the Latin datum, meaning a 'thing given'.
[13] Abdulrauf & Dube (2024).

bias or disenfranchisement.[14] It is common for conversations about systems for creating and sharing information to assume that such systems must inherently rely on digital data.

The ubiquity of this concession is unsurprising since digital infrastructures are established globally and largely outside deliberative democratic processes. For example, no local, national or international legislative body considered whether it would be a good idea to set up a system whereby people access news through personalized digital filters designed to maximize the likelihood that they will eventually buy something. Digital infrastructures are imposed primarily because of under- or unregulated corporate activity alongside opaque government procurement processes.[15] Many factors help to create the conditions in which the data-related features of infrastructures proliferate, typically with little political friction. None is more crucial than the lack of robust, and robustly enforced, rules about which public and private actors can do what with respect to data.[16]

Instances of political friction can generate political participation and change'.[17] Thanks to increasing public concern about corporate data practices that followed OpenAI's release of ChatGPT in late 2022 and a generative AI 'arms race', discussion around data governance issues is at an all-time high. Policy makers have historically taken up data governance in relation to the privacy, security and integrity of data, but there is strong political pressure now from within civil society to think about data governance as a lever for restructuring the markets in which technology companies operate. This is leading to efforts to protect people against infringements of their human rights, and also against concentrations of power and wealth that result in practices that are inconsistent with democracy.[18]

This attention gives us an opportunity to question the roles of digital data, data-dependent digital infrastructures, data markets and companies in the data business in the very formation and function of information ecosystems. Such questioning must be part of any democratic digital policy-making project, but any such project must also seek to preserve and promote the capacities of diverse communities to take up such questioning *outside* formal policy-making spaces. This questioning is necessary *not only for* democracy, but as democracy.[19]

Consistent with Arendt's view, what is at stake when the data and information about the world is structured by technologies that few understand, and even fewer control is not so much the ability to resist the manipulations of technologies (as important as that may be); it is the ability to think and deliberate with others about the meaning of information and of information systems in relation to the common good.[20] The issues here extend far beyond protecting and promoting a healthy and inclusive public sphere, because the data practices that undergird today's information ecosystems have profound social, economic and political implications (e.g., relating to environmental impacts or wealth distribution). To explore these issues, it is necessary to understand how corporate data monopolization impacts these systems.

3 Corporate Data Monopolization and Information Infrastructures

For people living in places with a highly developed digital infrastructure, it is almost impossible to live without creating a digital record of their lives. This is increasingly so when this infrastructure starts to become more accessible to those in the

[14] Chair (2024).
[15] Calo & Citron (2021); Colclough (2022); Crump (2016); Hardy & Williams (2008); Zuboff (2019).
[16] Cohen (2019).
[17] Gordon-Tapiero *et al.* (2023); Salehi *et al.* (2015).
[18] Doctorow & Giblin (2023); Mejias & Couldry (2024).
[19] Benson (2019); Chambers (2023), supported by the Economic and Social Research Council UK.
[20] See Mazzucato (2023) for one perspective on the 'common good' as distinct from the 'public good' concept.

Global Majority World. The publication of *The Age of Surveillance Capitalism* in 2019 coincided with a surge of attention to the ways that corporations track, record and analyze our online activities to predict and shape consumers' behavior.[21] This is the data that companies such as Google, X (formerly Twitter), Meta and Microsoft collect when people use their apps and services to create, find, consume and share information, but this is only a small part of a vast data surveillance landscape, a landscape that includes most government bureaucracies and most public and private entities that manage services and industries that are most crucial for the public.

When we do anything in a data economy, data is being produced about us: when we take public transport or drive on public roads, when we work at our jobs, when we open a bank account or apply for a credit card, when we have an interaction with a police officer or seek judicial intervention, when we apply for public benefits, when we rent or buy a home or sign up for electricity for that home, when we go to school, when we go to the doctor, when we interact in online spaces.[22] As the amount of data and the number of digital repositories grow exponentially, so do the networks and digital mechanisms for sharing and selling data. Government agencies are often unaware of who has access to the data they produce about their constituents,[23] although, depending on the context, there may be rules about what the collector of data can do with it, with whom they can share it, and what a third party can then do with it.

Existing data governance frameworks consist of 'a patchwork of national regulatory regimes, multilateral bodies, corporate policies, and multi-stakeholder organizations',[24] and these have not proven sufficient to protect most kinds of data from being acquired by large companies that use it to generate profit or amass power.[25] People are being comprehensively surveilled through data production, and the companies whose products and services shape information ecosystems are monopolizing this data. The political economy of the data infrastructure is one in which almost everyone is digitally surveilled, with the risks and burdens falling unequally on different groups and weighing most heavily on those who are already vulnerable, exploited, marginalized or targeted outside the digital context.[26] Those who are economically disadvantaged or subject to any form of group oppression are impacted disproportionately by the negative impacts of digital information ecosystems and by downstream misuses of data that companies commit within and for information systems, exemplified by data extraction without or with weak consent and by deploying algorithms biased in ways that benefit their economic performance. This is only compounded by the prevalence of mis- and disinformation.

The more companies achieve control of data, the more difficult it is to enact structural and systemic changes to address injustice and inequality in the digital era. There is a wide variety of corporate data practices that contribute to dysfunction and unfairness. Most involve two main types of monopolistic activity: monopolization of user data (i.e., all the data produced about us), which makes money for companies by converting information seekers into 'information products' offered for sale to advertisers; and monopolization of knowledge (i.e., data organized as usable insight) and information that makes money by converting data resources (including public data resources) into private assets. These pervasive forms of datafication give rise to numerous forms of digital dependency.

3.1 DATA MONOPOLIZATION AND DATA DEPENDENCY

Larry Page and Sergey Brin wrote an article in 1998 expressing concern about the ways that

[21] Abdulrauf & Dube (2024).

[22] On transportation, see Díaz & Levinson-Waldman (2020), supported by the Digital Industry Group Inc. (DIGI), an Australian not-for-profit industry association; on jobs, see Ajunwa *et al.* (2017); on banking, policing and court procedures, see Brayne (2020); on benefits, see Eubanks (2018); on homes and energy, see Harwell (2021); on schools, see Hooper *et al.* (2022); and on doctors, see Ledford (2019) and Corrales Compagnucci *et al.* (2022).

[23] Harwell (2019).

[24] LaForge & Gruver (2023).

[25] Mulligan & Godsiff (2023); The Majority Report w/ Sam Seder (2023); Zuboff (2019); see also Chapters 6, 7 and 8.

[26] Benjamin (2019); Browne (2015); Eubanks (2018); Fontes *et al.* (2022); Graham & Dittus (2022); Noble (2018); O'Neil (2016).

advertising revenue might affect the integrity of their newly launched internet search engine, Google: 'we believe the issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm'.[27] Three years later, after unsuccessful attempts to sell Google to other search companies, they began selling advertising based on user data. Between 2001 and 2003 Google's revenue increased 3,590 percent, from USD 19 million to USD 3.2 billion.[28] Four companies – Alphabet (Google), Meta (Facebook), Amazon, Microsoft – now largely control people's experience of using the internet to discover and share information. The logic of the advertising attention or 'eyeball' economy dictates the kinds of information a person can find or receive, whether that information relates to spring fashion trends, the next election, or the history of capitalism. In the Global North, these big tech companies also dominate advertising markets – Facebook and Google together control 70 percent of the market in the United States and over 65 percent in the United Kingdom.[29] It is the troves of data they collect about their users in all countries in which they operate that enables them to exert such market dominance. The practices of other digital platform companies are similar even if they command smaller market shares.

These big tech companies do not limit their data collection activities to the data that they themselves extract about the people using their digital products; they also buy or license data from other companies (and acquire data analytics companies) that gather data from a wide range of public and private sources.[30] They also scrape and aggregate massive amounts of data from every corner of the internet. And these companies are not transparent about their uses of the data they purchase from third parties or compile from public sources, using it for targeted advertising and product development – including the training of the algorithms that structure information flows through content moderation and curation, driving consumer activity. They sell their own data in ways that are extremely dangerous for democratic societies.[31] If, for example, hospital systems become dependent on a managed care algorithm owned by Microsoft, the company would have significant leverage over hospital decisions about how to deliver medical care, and could make it difficult for governments to limit their data practices and data hoarding that is required to train and maintain the technology undergirding the managed care algorithm.[32]

The fact that companies now take data from the internet without having to justify or compensate the data owners in any way is not the result of policy to affirmatively permit such activities. In many cases, companies determine what, if any, limits they will abide by, and they can change these at any time. Google's 2007 terms of service read:

> You give Google a perpetual, irrevocable, worldwide, royalty-free, and non-exclusive license to reproduce, adapt, modify, translate, publish, publicly perform, publicly display and distribute any Content which you submit, post or display on or through, the Services.[33]

In the absence of robust data governance, tech companies treat data as an exploitable resource and, following a playbook similar to the history of European colonialism, use that data to create conditions in which resisting their continued use of that data becomes both difficult and costly.[34] Big tech companies use their power to amass data to reinforce their advertising dominance, squeezing out competitors, and making it difficult to develop a product or service around an alternative set of data practices, or to use, test and scale up a framework for information sharing that does not depend on advertising.

[27] Brin & Page (1998), funded by DARPA and NASA, and Interval Research.
[28] Veliz (2021, p. 32).
[29] Doctorow & Giblin (2023).
[30] Savitz (2019); The Majority Report w/ Sam Seder (2023).
[31] See Biddle (2024), and, as in the case of Cambridge Analytica's acquisition of data for political targeting, see Briant (2021) and Dowling (2022), supported by Department of Defense, Australia.
[32] Tucker (2023).
[33] Mejias & Couldry (2024).
[34] Mejias & Couldry (2024).

Data monopolization is becoming more extreme as companies race to acquire the massive amounts of data needed to train algorithms to perform sophisticated classification tasks and predictive modelling. These technologies, marketed as 'AI', are extremely expensive to build because of the amount of data and computer-processing power required for training.[35] This means that only a small number of companies have the resources to compete in the development and training of data-intensive algorithms, and they are betting on an eventual payday that will justify their exorbitant initial investments.[36] The tech companies plan to integrate these technologies into their models for generating ad revenue (e.g., by using generative AI (GenAI) chatbots to mediate search activity), but it is unlikely that advertising revenue alone will yield profits that can justify the size of the bet that companies are placing on data-intensive algorithmic technology.

Every indication from big tech marketing and public relations documents is that their plan is to develop sophisticated industry-specific digital products that will offer to improve efficiency and reduce costs for companies operating within that industry, while creating 'path dependencies' that render client companies dependent on and thus 'locked in' to their algorithmic products (see Figure 4.1).[37] Bill Gates, whose company has invested USD 13 billion in OpenAI (that produced ChatGPT; see Figure 4.2), predicted that AI will redefine whole sectors of the economy and fundamentally change healthcare and education.[38]

Figure 4.1
Generative AI promotion

Source: AWS Amazon.com

Figure 4.2
Large Language models for efficiency

Source: WP Event Manager

In China the government is helping to facilitate a tech oligopoly, which, at least domestically, wields economic and political power that rivals that of the largest US-based tech companies. It is doing so by strategically cultivating its national technology champions – partly by banning foreign competitors and partly by using policy incentives to favor domestic firms.[39] As in the United States, a number of large tech companies has emerged in China that control digital platforms for social media, e-commerce, search and online payments.[40] The country's tech industry is also expanding its geopolitical influence through the Digital Silk Road (DSR) initiative.[41] This involves 'exporting' Chinese

[35] Mulligan & Godsiff (2023).
[36] Metz (2023); Novet (2023).
[37] OpenAI's website highlights a series of 'AI' products for sectors including healthcare and legal services (see https://openai.com/api). These companies are following the example of enterprise resource planning (ERP) system providers, such as SAP (Ven *et al.*, 2008); see also Ferràs-Hernández *et al.* (2023); Melih (2022); van der Vlist *et al.* (2024), supported by Dutch Research Council and German Research Foundation.
[38] Gates (2023).
[39] Tusikov (2021).
[40] Borgogno & Zangrandi (2024); Tusikov (2021); Wang (2023).
[41] Erie & Streinz (2021).

information and telecommunication goods to countries (especially in Africa and the Indo-Pacific) where digital infrastructure is in the early stages of development:

> … while it is true that China has consistently advocated for national autonomy over data governance issues, it is also trying to establish technological and infrastructural dependencies within the digital space of several countries. Technological dependences are established whenever the digital infrastructure relies on standards, software and hardware that cannot be maintained without active support from Chinese players.[42]

The problem with these data monopolization strategies is not just that a small number of companies control most of the digital data in the world or that they are using it primarily for self-enrichment.[43] It is also that weakly regulated corporate practices are what determined that digital data would be produced on such a scale and commercialized by default. This profoundly undemocratic, economic fiat succeeded in pre-empting meaningful political deliberation about rights to the ownership of digital data, what role data should have in the private and public sectors, how it should inform bureaucracy, and whether and in what contexts data production should be minimized or prohibited. Some political communities are having conversations about this now, but the terms of the debate are limited because of the entrenched data dependency globally across so many industries.[44]

The cost of having ceded so much ground to tech companies is especially high when it comes to information ecosystems. When it is assumed that contemporary systems for extracting and disseminating information must operate by reducing information to 'codable data' and using AI tools to determine how data should flow, the questions available for people to ask about what a good system should look like are rendered relatively one-dimensional. It may be possible for governments to compel tech companies to deliver information ecosystems that deploy algorithms that are more accurate, useful, inclusive or accessible. However, fixing the operational flaws of such algorithmic systems is ultimately a trivial problem compared with the problem for democratic society when it is not open to citizens, their representatives or the wider political community, including migrants and refugees, to contest the design and function, and even the existence, of these systems. It is this kind of thinking that involves *questioning the premise* – which is fundamental to political deliberation in democratic societies. Even the most sophisticated forms of AI do not produce systems that can question their own logics, that is, that can reflect on whether the 'learning' being done is the *right* learning for the problem at hand.[45]

3.2 BIG TECH MONOPOLIZATION

Working in parallel with the biggest tech companies in the Global North is a cohort of lesser known, but equally powerful, companies that make money by amassing and analyzing and then selling analyzed data sets to other companies and institutions. Examples of the most powerful companies in the data analytics world in Western countries are RELX, Thomson Reuters, and Experian. These companies horde and sell raw datasets, but also – and more importantly – information that has already been extracted from them.[46] RELX and Thomson Reuters, for example, have a duopoly with respect to legal information in the United States, owning the only two robust databases on the market for conducting legal research, both with high subscription fees. Experian is the dominant international data analytics company for financial information. It collects financial data from thousands of sources, analyzes and sells it, along with digital tools for integrating information, to businesses in more than 200 countries. There are smaller specialist data analytics companies that serve specific sectors, but

[42] Borgogno & Zangrandi (2024, p. 19).
[43] Cohen (2019); Melih (2022); Mulligan & Godsiff (2023); Sadowski (2019).
[44] Rankin (2023).
[45] Green (2022); Green & Kak (2021); Selbst *et al.* (2019), supported in part by NSF and Luminate (The Omidyar Group).
[46] Lamdan (2022).

the trend is towards large data analytics companies buying others to become behemoths that own and market digital data and information products across all markets.[47]

Some of the purchasers of these digital information products are other big tech companies, such as Google, and digital platforms, which incorporate the products into AI systems for search, newsfeed or ad placement. However, a large proportion of the customer base for data analytics companies is comprised of government agencies, financial institutions, law firms, universities, healthcare conglomerates and legacy media companies and news organizations. These are the entities that are typically able to pay for the products that data analytics companies sell, and they tend to function as economic, social and political gatekeepers within society. The result in many cases is that an enormous amount of information with profound public interest value is removed from the public sphere and reconstituted as the intellectual property of companies.[48] Any individual person who wants to access, for example, a newspaper article in the news archive owned by RELX containing '5 billion documents and records from over 35,000 sources of local and international news'[49] either has to be wealthy enough to afford a personal subscription, or be affiliated with an organization that has one. The consequence of placing so much information behind paywalls is often tragically concrete: 'doctors battling malaria outbreaks in Africa can't read reports about life-saving medications and measures. They can't afford to read past the articles' abstracts'.[50]

The situation is especially problematic when it comes to academic research, much of which is publicly funded by taxpayers and very little of which is available to taxpayers who do not belong to an elite institution.[51] Today, 'seventy-five percent of academic research is paywalled, and it usually costs around $30 to look at a single journal article'.[52] In addition to creating access-to-knowledge disparities for individuals based on institutional privilege, this creates disparities among institutions. Universities in the Global Majority World are less likely to have the resources to purchase subscription services within which companies like RELX trap the academic articles they own. For instance, 'in 2008, Harvard subscribed to 98,900 serials and Yale to 73,900. The best-funded research library in India, at the Indian Institute of Science, subscribed to 10,600. Several sub-Saharan African university libraries subscribed to zero, offering their patrons access to no conventional journals except those donated by publishers'.[53] Scientists from all over the world recruited 27 established institutions to try to access full-text paywalled articles in the field of ophthalmology. The results showed that at 15 of those institutions researchers could access less than half the articles. Those from institutions in wealthier countries (e.g., the United States and the Netherlands) were more likely to access most of the articles. Those at institutions in Pakistan and Ecuador were unable to access any of the articles.[54] This kind of control and dominance of the information ecosystems is illustrated, for example, by Elsevier's practices.

Controlling information products: Elsevier was founded in 1880 as a publisher of scientific and medical research. It is now owned by RELX and is the dominant player in a group of five companies that control access to academic research globally. Elsevier publishes over 500,000 academic articles annually in 2,500 journals, and its archives contain over 17 million documents. With its market control, Elsevier can charge

[47] Gautier & Lamesch (2021); Lamdan (2022); Larivière *et al.* (2015).

[48] Larivière *et al.* (2015), and see Chapter 3 for a discussion of copyright and AI systems. Some government agencies do provide open access to information with the costs of acquiring and curating it borne through taxation, but they increasingly outsource to services offered by private companies.

[49] LexisNexis (2024).

[50] The Majority Report w/ Sam Seder (2023, p. 53).

[51] Demeter (2019); Harvie *et al.* (2013); Nettle (2023); Puehringer *et al.* (2021).

[52] The Majority Report w/ Sam Seder (2023, p 53). Figures apply to the United States; charges are even higher in other parts of the world. This is changing, with many academic funders in the Global North mandating that researchers publish papers (and data where available) under open access rules. See EC (2016a); NSF (2023).

[53] Suber (2012, p. 30); Peters (2016).

[54] Boudry *et al.* (2019).

universities exorbitant fees for access to its journals, and package them in ways that boost the company's profit margins, rather than serving the needs of libraries and the people who use them. Elsevier makes thousands of dollars from journal articles that cost about USD 600 each to produce, generating a profit margin of 38 percent in 2023. This compares with the entire academic research industry's profit margins that hover around 30 percent, compared to Walmart's 3 percent and Toyota's 12 percent margin. After rebranding itself as an 'information analytics business', it began harvesting the data from its own content stores and using it to create digital products that do not serve the work of academics or researchers or students or librarians, but rather the money-making interests of research institutions. These products score and rank universities, journals and scholars according to prestige and influence metrics. They make predictions about which research projects will be successful. Some universities use these products to make hiring decisions, and academic funders use them to decide where to direct financial support.[55]

3.3 BUSINESS MODELS AND MIS- AND DISINFORMATION

A report for the Broadband Commission, supported by the International Telecommunication Union (ITU) and UNESCO, argues that platform business models make mis- and disinformation-based campaigns attractive.[56] A core driver behind these models is an 'economics of emotion' that depends on attention and incentivizes the creation of mis- and disinformation. Focusing on the 'politics of emotion', one study put it this way: 'as the design of the algorithms and interfaces of globally dominant social media platforms maximize emotional engagement, we regard social media as a primary site of datafied emotion worldwide'.[57]

'Digital influence mercenaries' exploit platform affordances on behalf of their clients, and mis- and disinformation is used to gain platform users' attention, transforming this into a commodity for sale to advertisers.[58] The introduction of the 'Like' button is a key step in the evolution of platform affordances that address the needs of influence mercenaries and platforms.[59] The 'Like' button gave Facebook a huge new source of valuable data about its users by tapping into their feelings while enhancing Facebook's personalization offering to advertisers. Facebook and other platforms are continually testing improved algorithms on its users for personalization.[60]

Almost all platforms have adopted variations of these 'vanity metrics', using them to algorithmically curate content posted by other users for the purposes of identifying specific users as recipients and as targets for advertisers.[61] How this content curation is performed varies among platforms. On TikTok, users' 'likes' are combined with their content views.[62] 4chan curates users' posts so that only the most 'liked' ones survive, which leads to the promotion of more extreme material.[63] Platforms such as YouTube reward content producers, which is argued to incentivize the creation of more extreme content. X (formerly Twitter) has adopted a variation of this policy.[64]

The question is whether these business models *inevitably* lead to platforms turning a blind eye to mis- and disinformation. After all, the reason the attention economy is key to platforms' financial viability is due to the advertising revenues that they generate on

[55] See Nicholson (2024); The Majority Report w/ Sam Seder (2023, p. 54).
[56] Bontcheva *et al.* (2020).
[57] Bakir & McStay (2022, p. 32).
[58] See Chapter 2 for further discussion of platform business models and for a definition of affordances.
[59] Bakir & McStay (2022).
[60] For example, by so-called 'A/B' testing, where the reactions to users of two different versions of a website are tested, often without the users being aware.
[61] Rogers (2018). This should not be confused with content moderation, which assesses whether a post is acceptable under a particular platform's rules.
[62] Benton (2022).
[63] Tuters & Hagen (2020), supported by the European Commission.
[64] Pequeño IV (2023).

the back of it. This kind of extreme content has been found to increase user engagement: the economics of emotion monetizes deception online, first, via a service contract with digital influence mercenaries to exploit platform affordances to achieve a client's strategic objectives, and second, by attracting user attention through deceptive content and selling this attention to advertisers.[65]

Having no content moderation policies (or policies that are not implemented) can be bad for business, as illustrated when advertisers terminate their business with platforms because their ads appeared alongside offensive posts such as hate speech, triggering 'brand boycotts'.[66] On balance, platform business models incentivize a very light-touch approach to content moderation and access to facilities such as their application programming interfaces (APIs).[67] X (formerly Twitter) pursues very permissive policies, allowing developers to use its APIs to create bots that automatically tweet. Facebook imposes tighter controls on the use of its APIs, but faces challenges in the management of the use of 'sock puppet' accounts (i.e., a false identity used for purposes of deception).[68]

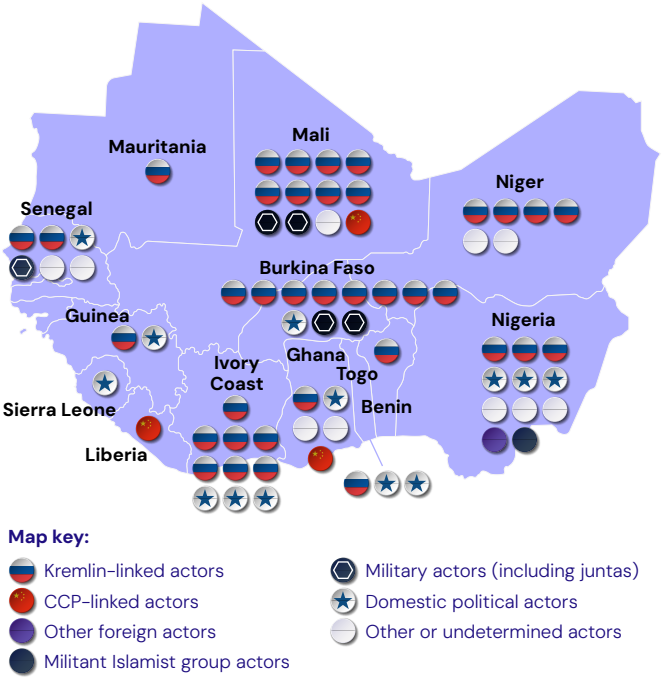
Comparing the business models of pre-digital broadcasting media, partisan media and digital media platforms reveals qualitatively different forms of mis- and disinformation. In the pre-digital media era, offending the audience was often considered bad for business, which encouraged the in some countries encouraged the presentation of news and information in ways that aligned with the beliefs and values of the majority, and with fewer tendencies towards what is today described as polarization. Regulatory changes in the United States created opportunities for the emergence of partisan media, which led to a growth in confrontational narratives and the promotion of minority viewpoints, while partisan or state controlled media were common in other parts of the world.[69] Online digital platforms have developed more sophisticated ways of using AI systems to create platform affordances that enable the exploitation of the capacity of controversial content to capture user engagement. These platform affordances are also being exploited by a growing army of influencers who are building a following and channel content that is designed to sustain their followers' engagement, which they are then able to monetize.[70]

A market-shaping approach is helpful for revealing how 'market-makers bring markets into existence through their day-to-day practices, and how their goal of generating viral content – and "clickbait" – incentivizes the circulation of "controversial claims, adversarial narratives and deceptive content"'.[71] Thus, mis- and disinformation are an 'expected outcome, not breakage', of the platformized media market: far from being evidence of a dysfunctional business model, it is the outcome that is expected given these business models.

Political campaigns world-wide are increasingly data-driven as the platforms' capacities to deliver targeted advertising become more sophisticated. The political sphere is an area where platform policies regarding transparency are a particularly important concern.[72] The Africa Center for Strategic Studies reported that mis- and disinformation campaigns increased nearly fourfold from 2022 to 2024, with a total of 189, and nearly 60 percent of these campaigns were sponsored by foreign states, with Russia, China and the United Emirates being prominent (see Figure 4.3).[73]

[65] Bakir & McStay (2022).
[66] Zhu (2022), supported by Finnish National Agency for Education.
[67] Zammit et al (2021), supported by the Erasmus+ Strategic Partnership Program; see also Gorwa & Guilbeault (2020). APIs provide a means to programmatically interface with platform data. One example of a problem arising from this is the harvesting of millions of user profiles that are then used for targeted political advertising (Hinds *et al.*, 2020, supported in part by the Economic and Social Research Council, UK).
[68] Wikipedia (2024).
[69] See Chapter 2 for an extensive examination of legacy and online news media.
[70] Diaz Ruiz (2023).
[71] Diaz Ruiz (2023, p. 1).
[72] Mehta & Erickson (2022).
[73] Africa Center for Strategic Studies (2024).

Figure 4.3
Map of mis- and disinformation campaigns in West Africa, and state sponsorship



Source: Africa Center for Strategic Studies (2024)

Responding to concerns about political advertising's lack of transparency, in 2018, Google, Facebook and X (formerly Twitter) established political ad archives, including information about advertisers. However, various factors, including doubts about the capacity of citizens to find and understand data and financial incentives, work against the delivery of transparency in an effective and meaningful way.[74]

The United Nations argues that 'digital platforms should move away from business models that prioritize engagement above human rights, privacy and safety'.[75] A report commissioned by NATO's Strategic Communications Centre of Excellence remarks that 'buying manipulation remains cheap' and 'the gap between countering inauthentic engagement and platform reporting is widening' as platforms focus more on limiting the reach and impact of messaging and less on denying access to commercial information manipulators.[76] Others observe that the cost of buying manipulation over time is stable,[77] raising the question of how it could be made more costly for major generators of mis- and disinformation, and what impact this might then have.

Those who are intent on spreading mis- and disinformation exploit the platforms' business models and encourage their complicity in campaign propagation.[78] For example, platforms provide a target for 'cyber troop' activity (i.e., government or political party actors tasked with manipulating public opinion online), with activity identified in at least 81 countries.[79] The data economy fosters a highly competitive labor market in datafication where people persuade themselves that if they do not take on work, others will. Anonymity is afforded to online laborers when they participate in datafication work.[80] In one experiment, 87 percent of participants were found to be willing to accept jobs involving the creation of mis- and disinformation.[81]

Existing data governance rules permit and even foster the amplification of mis- and disinformation through coordinated influence operations as, for example, in Venezuela. A study there revealed a range of influencer motivations, organizations, technical systems, adversaries and strategies, including recruiting and paying influencers with campaigns that were organized through hierarchies and decentralized operations. Those propagating mis- and disinformation learned continuously how to evade any defenses the platform (in this case Twitter) created.[82] The powerful big tech companies operate in ways that are counterproductive to efforts to tackle mis- and disinformation.

[74] Mehta & Erickson (2022).
[75] UN (2023a, p. 23).
[76] Fredheim *et al.* (2023, p. 3).
[77] Bradshaw *et al.* (2021), supported by the European Research Council (ERC), Adessium Foundation, Civitates Initiative, Ford Foundation, Hewlett Foundation, Luminate, Newmark Philanthropies and Open Society Foundations.
[78] Posetti & Bontcheva (2020).
[79] Bradshaw & Howard (2019) supported by European Research Council (ERC), Adessium Foundation, Hewlett Foundation and Luminate.
[80] Shleifer (2004).
[81] Cohn *et al.* (2022).
[82] Recabbaren *et al.* (2023), reporting on a semi-structured interview-based study with 19 participants. Interviews focused on: (1) incentives to contribute; (2) organizational structure; (3) resources, capabilities and limitations; (4) strategies employed; (5) operations they had participated in; (6) perception of disinformation in influence operations; (7) perception of the robustness of Twitter's defenses against influence operations activities; and (8) strategies to evade and recover from detection.

Other studies of online activity that is antithetical to healthy information ecosystems emphasize that it is crucial to examine the political economy of this activity – and not just the individual actors. For example, a study of mis- and disinformation campaigns in the Philippines and Indonesia (countries with high levels of social media activity) concluded that legacy media's history of ownership and political collusions in postcolonial societies makes them vulnerable to narratives about 'bias' and 'bigotry'. In this case, research highlighted 'the broader (Western) discourse) that has positioned these two countries as examples of Global Majority contexts where social media have "ruined democracy", insofar as masses of voters are assumed to have been duped by digital disinformation campaigns'.[83] It is a major problem when research focuses on individuals instead of on the political economy of infrastructures and data monetization, which enables platform complicity in encouraging the villainy of mis- and disinformation actors.[84] This confirms the need for creative approaches to data governance and for democratic decision making about who should be able to make use of data.

4 Towards Democratic Data Governance

The two varieties of corporate data monopolization (to control capital and access to knowledge) work in tandem to shape information ecosystems. The tech companies that are producing and using data for directing commercial behavior are maintaining an impoverished public sphere, which serves as the default digital space in which people discover and share information.[85] Another set of companies is capturing and cordoning off from public access rich sets of data that contain usable insights. The result is a radical difference between how an information ecosystem is experienced when operating in frameworks of institutional privilege and when operating outside such frameworks.

Where data is accessible through data repositories and networks that are designed to help people easily use this data to extract information to build knowledge, people can see themselves as readers, thinkers and participants in discourse. When most information available today is a byproduct of corporate data practices that primarily aim to satisfy goals unrelated to the social project of knowledge production, people are forced to fight a system designed to treat knowledge seeking as a consumer activity.[86] This makes it extremely difficult for people inhabiting a digital space to have experiences of shared inquiry. This dichotomy contributes to the spread of mis- and disinformation: 'in a world where scholarly research is paywalled, it's free to hop on YouTube to watch white supremacists spread racist theories about IQ and race. But, to read a scholarly article refuting the racist YouTubers' baseless claims with well-researched facts, the charge is USD 37.50 to overcome Sage Publishing's paywall'.[87]
Open access does not resolve all the constraints on access to online reliable information, but it does at least reduce the cost barrier where a digital infrastructure is in place.

The comprehensive data surveillance that underlies digital infrastructure, and the way that corporations aim to monopolize surveillance data to control who has access to what kinds of information, means that the role of data in information ecosystems should not be considered solely in terms of individual impacts. The inadequacy of conceptualizing the harms of the data economy in individualistic

[83] Ong & Tapsell (2022, p. 252), supported by the Department of Foreign Affairs and Trade, Australia.
[84] Ong & Tapsell (2022, p. 265), footnote in quote omitted.
[85] Franks (2021).
[86] Couldry & Mejias (2019); Magalhães & Couldry (2021); Schoon *et al.* (2020); West (2019).
[87] The Majority Report w/ Sam Seder (2023, p. 74).

terms is easy to see when thinking about the way that monopolistic corporate behavior makes access to high-quality information contingent on social and economic privilege. However, a more multidimensional account of data harms is necessary to make sense of the asymmetrical structure of the markets the major tech companies are using data to create.

> The data collection practices of the most powerful technology companies are aimed primarily at deriving (and producing) population level insights regarding how data subjects relate to others, not individual insights specific to the data subject. These insights can then be applied to all individuals (not just the data subject) that share these population features.[88]

The most prominent approaches to data governance (including AI governance) tend to focus on protecting security (individual and/or state), property and dignity/autonomy. Robust enforcement might improve outcomes for individuals and communities in highly datafied societies, but these frameworks fail in providing a framework for contesting datafication itself.[89] A primary motivation for companies to produce and aggregate massive amounts of data is to make predictions about group membership, group characteristics and behaviors that facilitate targeting for economic or political purposes, often to enhance targeted marketing of goods and services as well as for personalizing content.

While some existing or proposed frameworks for data governance address problems of improper economic or political influence, they do not take up the underlying data practices that make such targeting possible. They do not take account of the way data practices that strive to shape individual behavior according to predictions of AI tools about group membership can produce or entrench social injustices (such as wealth disparities or racial oppression).[90] Population level data injustice is related to, but conceptually distinct from, individual harms that people suffer as a result of facial recognition algorithms trained on data that reflects racial bias.[91] A neglected problem that arises with training AI systems at the population level is that the burdens of data production are borne disproportionately by certain groups. To understand what this means concretely, consider the case of a company called Fog Data Science.

Population level data injustice: Fog Data Science is a company based in the United States founded by two former Department of Homeland Security officials. Its main product is a digital tracking program called Fog Reveal, which it sells primarily to law enforcement agencies. Police departments that subscribe to Fog Reveal have access to a database containing billions of records from 250 million mobile devices, and can conduct a variety of different searches (including search by device ID). Based on these searches, the police can develop a 'pattern of life analysis' – a profile of individual habits based on long-term behavioral data. Fog Data Science built and maintains the Fog Reveal database by buying domestic and international location data from data brokers, which originates from over 700 smartphone apps using a mechanism called an 'ad ID'.

The ad ID was created as a way for advertisers to personalize offers for mobile device users. It is a random string of numbers and letters that attaches to the data that smartphone apps generate about users – bundles of data can include private information (year of birth, gender, search terms used and location). Most mobile device users do not know about ad IDs or how they work, and those who are aware that their apps are recording data about their movements rarely have a way of knowing that this data is being purchased by data brokers and

[88] Viljoen (2021, p. 577).
[89] Datafication is defined in Section 2, Chapter 1. AI system governance through legislative approaches is discussed in Section 4.4, Chapter 6 and Section 3.1, Chapter 7.
[90] Viljoen (2021).
[91] Mayson (2018).

sold to the police for surveillance purposes. If an individual is aware that there is a trade-off involved in using an app, they think of it as a trade-off of their privacy for their convenience. They have no way to predict downstream uses of their data or to orient their behavior ethically with an awareness of potential downstream harms.

The fact that ad ID is being used non-transparently for policing creates risks that disproportionately impact some groups more than others. In the United States, a middle-aged White woman is much less likely to be targeted by police using Fog Reveal than a young Black man, so Black men will disproportionately experience the harms not only of Fog Reveal, but of all the companies in the chain through which Fog Reveal obtains data.[92]

This illustrates a pervasive problem in a political economy where average levels of awareness of, and tolerance for, privacy invasion set the limits of data practices, with drastically differential real-world consequences for different groups of people. Democratic data governance requires political structures that make it possible for communities to grapple with how data production and information extraction impacts on the distribution of power among people and those entities engaging in these practices, who are already differently situated – socially, culturally and economically:

> The status quo of data governance law, as well as prominent proposals for its reform … attempt to reduce legal interests in information to individualist claims subject to individualist remedies, which are structurally incapable of representing the interests and effects of data production's population-level aims. This in turn allows significant forms of social informational harm to go unrepresented and unaddressed in how the law governs data collection, processing, and use.[93]

Given the global reach of big tech companies, and the ease with which data is created, shared, transferred and copied,[94] data governance interventions need to consider how population level effects of data practices manifest within, but also across, political boundaries.

5 Chapter Summary

This chapter has examined how digital data, data-dependent digital infrastructures, data markets and companies in the data business produce inequalities. It has focused on how powerful actors within social, economic and political systems determine what data is produced and how it is produced. Data aggregation techniques and the AI systems embedded in digital platforms and services are changing how people build, share and receive information and knowledge. This chapter examined research on how the business models of big tech companies contribute to the production of mis- and disinformation by creating incentives for individuals to engage in information production of this kind.

These models lead to the dependence of individuals and industry sectors on the technologies and services provided by big tech companies in data-intensive economies. The research synthesis was informed by a political economy approach which focuses on struggles to govern data practices to be consistent with people's rights and interests. This chapter has shown how the monopolistic power and data governance practices favored by big tech companies succeed in pre-empting meaningful political deliberation about issues such as rights to data ownership, what role data should have in the private and public sectors, and in what contexts data production should be minimized or prohibited.

It has emphasized that combating mis- and disinformation is a collective endeavor. It requires concerted action from governments, platform providers, civil society and political entities to question

[92] Cyphers (2022); Greenberg (2022); Turow *et al.* (2023).
[93] Viljoen (2021, p. 578).
[94] Quintais *et al.* (2023).

the fairness of technology development and data uses that have unfair discriminatory consequences or do not foster information ecosystems that uphold the integrity of democratic processes.

The synthesis of research in this chapter shows that:

- Dependence on data-intensive algorithmic products, marketed as 'AI', is growing, posing significant risks to democracy. This is because when data and information are structured in ways that few understand or have control over, this affects their abilities to resist manipulations and to think and deliberate with others about the common good.

- The monopolization of data (i.e., data organized as usable insight or knowledge) occurs by converting data resources (including public data resources) into private assets. People are surveilled for data and the big tech companies do not limit their data collection to the data they extract. They buy or license data from other companies (and acquire data analytics companies) that gather or process data. Other less well known, but similarly powerful, companies also participate by amassing, analyzing and then selling data sets to other companies and institutions.

- Data governance legislation and frameworks are sufficiently permissive to foster the amplification of mis- and disinformation. These governance arrangements mean that companies and their infrastructures are creating de facto data governance frameworks that are inconsistent with data justice, and these frameworks have become normalized.

- Understanding the role of data and machine learning technologies in information ecosystems requires a multidimensional analysis of data harms that is informed by how global data dependency is becoming entrenched – that is, it must go beyond the study of impacts on individuals to focus on the political economy of power relationships and the asymmetries they produce.

Research is needed:

- To investigate the tension between the benefits of building out network infrastructures and promoting the use of AI systems in countries in the Global Majority World where internet access is absent or very limited. Doing so risks entrenching the problems experienced in higher-income, data-intensive economies with their advanced digital infrastructures and claims to robust data governance regimes.

- To assess whether AI systems are developing in ways that are counterproductive to efforts (technical or otherwise) to tackle mis- and disinformation by investigating and exposing how big tech business models make them attractive targets for mis- and disinformation campaigns and encourage their complicity in such campaigns.

- To study how online labor markets incentivize the production of mis- and disinformation and the efficacy of steps that could be taken to discourage this.

- To investigate how extractive data production has harmful consequences for people's daily lives, with a focus on the replication and exacerbation of inequalities and injustices.

- To examine data governance frameworks devised in countries in the Global Majority World where they are still emerging or have only recently been put in place, in order to understand what strategies are available to resist the power of big tech companies.

References

- Aaronson, S. A. (2021). Data is Different, Policymakers Should Pay Attention to Its Governance. In M. Burri (Ed.), *Big Data and Global Trade Law* (pp. 340-360). Cambridge University Press.
- Abdulrauf, L. A., & Dube, H. (Eds) (2024). *Data Privacy Law in Africa: Emerging Perspectives*. Pretoria University Law Press.
- Africa Center for Strategic Studies. (2024). *Mapping a Surge of Disinformation in Africa*. https://africacenter.org/spotlight/mapping-a-surge-of-disinformation-in-africa
- Aguerre, C., Campbell-Verduyn, M., & Scholte, J. A. (Eds) (2024). *Global Digital Data Governance: Polycentric Perspectives*. Routledge.
- Ajunwa, I., Crawford, K., & Schultz, J. (2017). Limitless worker surveillance. *Berkely Law*, 105, 735-776. https://doi.org/10.15779/Z38BR8MF94
- Arendt, H. ([1958] 1998). *The Human Condition*. University of Chicago Press.
- Bakir, V., & McStay, A. (2022). Core Incubators of False Information Online. In V. Bakir & A. McStay, Optimising Emotions, Incubating Falsehoods: How to Protect the Global Civic Body from Disinformation and Misinformation (pp. 29-52). Springer International Publishing.
- Barba-Kay, A. (2023). *A Web of Our Own Making: The Nature of Digital Formation*. Columbia University Press.
- Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons.
- Benson, J. (2019). Deliberative democracy and the problem of tacit knowledge. *Politics, Philosophy & Economics*, 18(1), 76-97. https://doi.org/10.1177/1470594X18782086
- Benton, J. (2022). Facebook looks ready to divorce the news industry, and I doubt couples counseling will help. NiemanLab, 16 June. www.niemanlab.org/2022/06/facebook-looks-ready-to-divorce-the-news-industry-and-i-doubt-couples-counseling-will-help
- Biddle, S. (2024). Elon Musk fought government surveillance – While profiting off government surveillance. *The Intercept*, 25 March. https://theintercept.com/2024/03/25/elon-musk-x-dataminr-surveillance-privacy
- Bontcheva, K., Posetti, J., Teyssou, D., Meyer, T., *et al.* (2020). *Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression*. Working Group Report. Broadband Commission for Sustainable Development. www.broadbandcommission.org/publication/balancing-act-countering-digital-disinformation
- Borgogno, O., & Zangrandi, M. S. (2024). Chinese data governance and trade policy: From cyber sovereignty to the quest for digital hegemony? *Journal of Contemporary China*, 33(148), 578-602. www.tandfonline.com/doi/full/10.1080/10670564.2023.2299961
- Boudry, C., Alvarez-Muñoz, P., Arencibia-Jorge, R., Ayena, D., Brouwer, N. J., Chaudhuri, Z., *et al.* (2019). Worldwide inequality in access to full text scientific articles: The example of ophthalmology. *PeerJ*, 7, 1-18. https://doi.org/10.7717/peerj.7850
- Bradshaw, S., & Howard, P. N. (2019). *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation*. Oxford Internet Institute and University of Oxford. www.oii.ox.ac.uk/news-events/reports/the-global-disinformation-order-2019-global-inventory-of-organised-social-media-manipulation
- Bradshaw, S., Bailey, H., & Howard, P. N. (2021). *Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation*. Programme on Democracy and & Technology. Oxford Internet Institute and University of Oxford.
- Brayne, S. (2020). *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford University Press.
- Briant, E. L. (2021). *Propaganda Machine: The Hidden Story of Cambridge Analytica and the Digital Influence Industry*. Bloomsbury Publishing PLC.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems, 30(1-7), 107-117. https://doi.org/10.1016/S0169-7552(98)00110-X
- Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness*. Duke University Press.
- Calo, R., & Citron, D. K. (2021). The automated administrative state: A crisis of legitimacy. Emory Law Journal, 70(4), 799-844. https://scholarlycommons.law.emory.edu/elj/vol70/iss4/1
- Chair, C. (2024). Gendered Digital Inequalities: How Do We Ensure Gender Transformative Law and Practice in the Age of Artificial Intelligence in Africa? In L. A. Abdulrauf & H. Dube (Eds), *Data Privacy Law in Africa: Emerging Perspectives* (pp. 207-238). Pretoria University Law Press.
- Chambers, S. (2023). Deliberative democracy and the digital public sphere: Asymmetrical fragmentation as a political not a technological problem. Constellations, 30(1), 61-68. https://doi.org/10.1111/1467-8675.12662
- Cohen, J. E. (2019). *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford University Press.
- Cohen, J. E. (2023). Infrastructuring the digital public sphere. *Yale Journal of Law and Technology*, 25(1), 1-40. https://law.yale.edu/sites/default/files/area/center/isp/documents/cohen_julie_-_infrastructuring_the_digital_public_sphere.01.pdf
- Cohn, A., Stoop, J., & Rahman, H. A. (2022). *Disinformation for Hire: Examining the Production of False COVID-19 Information*. Tinbergen Institute Discussion Paper, TI 2022-086/II. https://papers.tinbergen.nl/22086.pdf
- Colclough, C. (2022). Reshaping the digitization of public services. *New England Journal of Public Policy*, 34(1), 1-14. https://scholarworks.umb.edu/nejpp/vol34/iss1/9

- Corrales Compagnucci, M., Wilson, M. L., Fenwick, M., Forgó, N., & Bärnighausen, T. (Eds) (2022). *AI in eHealth: Human Autonomy, Data Governance and Privacy in Healthcare*. Cambridge University Press.
- Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media*, 20(4), 336-349. https://doi.org/10.1177/1527476418796632
- Crump, C. (2016). Surveillance policy making by procurement. Washington Law Review, 91(4), 1595-1662. https://doi.org/10.2139/ssrn.2737006
- Cyphers, B. (2022). Inside Fog Data Science, the secretive company selling mass surveillance to local police. Electronic Frontier Foundation, 31 August. www.eff.org/deeplinks/2022/08/inside-fog-data-science-secretive-company-selling-mass-surveillance-local-police
- de-Lima Santos, M.-F. (2023). The entanglements between data journalism, collaboration and business models: A systematic literature review. *Digital Journalism*, 12(2), 256-281. https://doi.org/10.1080/21670811.2023.2247449
- Demeter, M. (2019). The winner takes it all: International inequality in communication and media studies today. *Journalism & Mass Communication Quarterly*, 96(1), 37-59. https://doi.org/10.1177/1077699018792270
- Díaz, A., & Levinson-Waldman, R. (2020). *Automatic License Plate Readers: Legal Status and Policy Recommendations for Law Enforcement Use*. Brennan Center for Justice. www.brennancenter.org/our-work/research-reports/automatic-license-plate-readers-legal-status-and-policy-recommendations
- Diaz Ruiz, C. (2023). Disinformation on digital media platforms: A market-shaping approach. New Media & Society, Online First, 1-24. https://doi.org/10.1177/14614448231207644
- Doctorow, C., & Giblin, R. (2023). *Chokepoint Capitalism: How Big Tech and Big Content Captured Creative Labor Markets and How We'll Win Them Back*. Beacon Press.
- Dowling, M.-E. (2022). Cyber information operations: Cambridge Analytica's challenge to democratic legitimacy. Journal of Cyber Policy, 7(2), 230-248. https://doi.org/10.1080/23738871.2022.2081089
- EC (European Commission). (2016a). *Background Note on Open Access to Scientific Publications and Open Research Data*. https://research-and-innovation.ec.europa.eu/document/download/4bd9ef8e-0101-457d-9fc5-1096c4e8f6f0_en?filename=ec_rtd_background-note-open-access.pdf
- Erie, M. S., & Streinz, T. (2021). The Beijing effect: China's 'digital silk road' as transnational data governance. *New York University Journal of International Law and Politics*, 54(1), 1-25. https://ssrn.com/abstract=3810256
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St Martin's Publishing Group.
- Ferràs-Hernández, X., Nylund, P. A., & Brem, A. (2023). The emergence of dominant designs in Artificial Intelligence. California Management Review, 65(3), 73-91. https://doi.org/10.1177/00081256231164362
- Fontes, C., Hohma, E., Corrigan, C. C., & Lürge, C. (2022). AI-powered public surveillance systems: Why we (might) need them and how we want them. *Technology in Society*, 71(2022), 1-12. https://doi.org/10.1016/j.techsoc.2022.102137
- Franks, M. A. (2021). Beyond the public square: Imagining digital democracy. Yale Law Journal, 131(2021-22), 427-453. www.yalelawjournal.org/forum/beyond-the-public-square-imagining-digital-democracy
- Fredheim, R., Bay, S., Dek, A., Stolze, M., & Haiduchyk, T. (2023). *Social Media Manipulation 2022/2023: Assessing the Ability of Social Media Companies to Combat Platform Manipulation*. NATO Strategic Communciations Centre of Excellence. https://stratcomcoe.org/publications/social-media-manipulation-20222023-assessing-the-ability-of-social-media-companies-to-combat-platform-manipulation/272
- Gates, B. (2023). The age of AI has begun. GatesNotes Blog, 21 March. www.gatesnotes.com/The-Age-of-AI-Has-Begun
- Gautier, A., & Lamesch, J. (2021). Mergers in the digital economy. Information Economics and Policy, 54(2021), 1-15. https://doi.org/10.1016/j.infoecopol.2020.100890
- Gordon-Tapiero, A., Ohm, P., & Ramaswami, A. (2023). Fact and Friction: A Case Study in the Fight Against False News. UC Davis Law Review, 57, November. https://lawreview.law.ucdavis.edu/archives/57/1/fact-and-friction-case-study-fight-against-false-news
- Gorwa, R., & Guilbeault, D. (2020). Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet*, 12(2), 225-248. https://doi.org/10.1002/poi3.184
- Graham, M., & Dittus, M. (2022). Geographies of Digital Exclusion: Data and Inequality. Pluto Press.
- Green, B. (2022). Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. *Philosophy & Technology*, 35(4), 1-32. https://doi.org/10.1007/s13347-022-00584-6
- Green, B., & Kak, A. (2021). The false comfort of human oversight as an antidote to AI harm. Slate, 15 June. https://ssrn.com/abstract=4046163
- Greenberg, W. (2022). How ad tech became cop spy tech. Electronic Frontier Foundation, 31 August. www.eff.org/deeplinks/2022/08/how-ad-tech-became-cop-spy-tech
- Hardy, C., & Williams, S. P. (2008). E-government policy and practice: A theoretical and empirical exploration of public E-procurement. *Government Information Quarterly*, 25(2), 155-180. https://doi.org/10.1016/j.giq.2007.02.003
- Harvie, D., Lightfoot, G., Lilley, S., & Weir, K. (2013). Publisher, be damned! From price gouging to the open road. *Prometheus*, 31(3), 229-239. https://doi.org/10.1080/08109028.2014.891710

- Harwell, D. (2019). FBI, ICE find state driver's license photos are a gold mine for facial-recognition searches. *The Washington Post*, 7 July. www.washingtonpost.com/technology/2019/07/07/fbi-ice-find-state-drivers-license-photos-are-gold-mine-facial-recognition-searches
- Harwell, D. (2021). ICE investigators used a private utility database covering millions to pursue immigration violations. *The Washington Post*, 26 February. www.washingtonpost.com/technology/2021/02/26/ice-private-utility-data
- Herman, E. S., & Chomsky, N. (1989). Manufacturing Consent: The Political Economy of the Mass Media. Penguin.
- Hinds, J., Williams, E. J., & Joinson, A. N. (2020). 'It wouldn't happen to me': Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*, 143(2020), 1-55. https://doi.org/10.1016/j.ijhcs.2020.102498
- Hooper, L., Livingstone, S., & Pothong, K. (2022). *Problems with Data Governance in UK Schools: The Cases of Google Classroom and Class Dojo*. Digital Futures Commission, 5Rights Foundation. https://digitalfuturescommission.org.uk/wp-content/uploads/2022/08/Problems-with-data-governance-in-UK-schools.pdf
- LaForge, G., & Gruver, P. (2023). *Governing the Digital Future*. Planetary Politics, October. https://core.ac.uk/download/588303818.pdf
- Lamdan, S. (2022). *Data Cartels: The Companies that Control and Monopolize Our Information*. Stanford University Press.
- Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. PLOS ONE, 10(6), 1-15. https://doi.org/10.1371/journal.pone.0127502
- Ledford, H. (2019). Millions of black people affected by racial bias in health-care algorithms. Nature, 574(7780), 608-609. https://doi.org/10.1038/d41586-019-03228-6
- LexisNexis. (2024). Sources of local and international news and business information. www.lexisnexis.ca/en-ca/products/nexis.page
- Local Progress. (2024). Creating Traffic Safety: A Policy Memo for Local Elected Leaders. 9 January. https://localprogress.org/resource/creating-traffic-safety-a-policy-memo-for-local-elected-leaders
- Magalhães, J. C., & Couldry, N. (2021). Giving by taking away: Big tech, data colonialism, and the reconfiguration of social good. *International Journal of Communication*, 15(2021), 343-362. https://ijoc.org/index.php/ijoc/article/view/15995
- Mayson, S. G. (2018). Bias in, bias out. The Yale Law Journal, 2393, 2218-2300. https://scholarship.law.upenn.edu/faculty_scholarship/2393
- Mazzucato, M. (2023). *Governing the Economics of the Common Good: From Correcting Market Failures to Shaping Collective Goals*. Working Paper 2023/08. Institute for Innovation and Public Purpose. www.ucl.ac.uk/bartlett/public-purpose/sites/bartlett_public_purpose/files/mazzucato_m._2023._governing_the_economics_of_the_common_good_from_correcting_market_failures_to_shaping_collective_goals.pdf
- McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. McGraw-Hill.
- Mehta, S., & Erickson, K. (2022). Can online political targeting be rendered transparent? Prospects for campaign oversight using the Facebook Ad Library. Internet Policy Review, 11(1), 1-32. https://doi.org/https://doi.org/10.14763/2022.1.1648
- Mejias, U. A., & Couldry, N. (2024). *Data Grab: The New Colonialism of Big Tech and How to Fight Back*. W.H. Allen.
- Melih, Y. (2022). New geographies of platform capitalism: The case of digital monopolization in Turkey. Big Data & Society, 9(2), 1-14. https://doi.org/10.1177/20539517221124585
- Metz, C. (2023). The ChatGPT king isn't worried, but he knows you might be. The New York Times, 31 March. www.nytimes.com/2023/03/31/technology/sam-altman-open-ai-chatgpt.html
- Morozov, E. (2013). *To Save Everything, Click Here: The Folly of Technological Solutionism*. Public Affairs.
- Mulligan, C. E. A., & Godsiff, P. (2023). Datalism and data monopolies in the era of AI: A research agenda. arXiv:2307.08049v1. https://doi.org/10.48550/arXiv.2307.08049
- Nettle, D. (2023). The political economy of scientific publishing, and the promise of diamond open access. 14 June. www.danielnettle.org.uk/2023/06/14/the-political-economy-of-scientific-publishing-and-the-promise-of-diamond-open-access
- Nicholson, C. (2024). Elsevier parent reports 10% hike in profits for 2023. *Research Professional News*, 15 February. www.researchprofessionalnews.com/rr-news-europe-infrastructure-2024-2-elsevier-parent-reports-10-hike-in-profits-for-2023
- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.
- Novet, J. (2023). Microsoft's $13 billion bet on OpenAI carries huge potential along with plenty of uncertainty. CNBC, 8 April. www.cnbc.com/2023/04/08/microsofts-complex-bet-on-openai-brings-potential-and-uncertainty.html
- NSF (National Science Foundation). (2023). *NSF Public Access Plan 2.0: Ensuring Open, Immediate and Equitable Access to National Science Foundation Funded Research*. www.nsf.gov/pubs/2023/nsf23104/nsf23104.pdf
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Harvard University Press.
- OECD (Organisation for Economic Co-operation and Development) (2022a). Going Digital to Advance Data Governance for Growth and Well-Being. https://doi.org/10.1787/e3d783b0-en
- 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

- Padovani, C., Wavre, V., Hintz, A., Goggin, G., & Iosifidis, P. (Eds) (2024). Global Communication Governance at the Crossroads. Palgrave Macmillan.
- Pequeño IV, A. (2023). Twitter starts paying creators tens of thousands – Amid intensifying competition with Threads. *Forbes*, 13 July. www.forbes.com/sites/antoniopequenoiv/2023/07/13/twitter-starts-paying-creators-tens-of-thousands---amid-intensifying-competition-with-threads
- Peters, J. (2016). Why is it so expensive to read academic research? *Slate*, 5 April. https://slate.com/technology/2016/04/the-lawsuit-against-sci-hub-begs-the-question-why-are-academic-journals-so-expensive-anyway.html
- Posetti, J., & Bontcheva, K. (2020). *Disinfodemic: Deciphering COVID-19 Disinformation*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000374416
- Puehringer, S., Rath, J., & Griesebner, T. (2021). The political economy of academic publishing. PLOS ONE, 16(6), 1-21. https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0253226&type=printable
- Quintais, J. P., De Gregorio, G., & Magalhães, J. C. (2023). How platforms govern users' copyright-protected content: Exploring the power of private ordering and its implications. *Computer Law & Security Review*, 48(2023), 1-25. https://doi.org/10.1016/j.clsr.2023.105792
- Rankin, J. (2023). MEPs launch site for EU officials to report 'shady lobbying' by big tech. *The Guardian*, 2 February. https://amp.theguardian.com/world/2023/feb/02/meps-launch-site-for-eu-officials-to-report-shady-lobbying
- Recabarren, R., Carbunar, B., Hernandez, N., & Shafin, A. A. (2023, 2023). Strategies and Vulnerabilities of Participants in Venezuelan Influence Operations. 32nd USENIX Security Symposium. www.usenix.org/conference/usenixsecurity23/presentation/recabarren
- Rogers, R. (2018). Digital traces in context| Otherwise engaged: Social media from vanity metrics to critical analytics. *International Journal of Communication*, 12(2018), 23. https://ijoc.org/index.php/ijoc/article/view/6407
- Rosenberg, D. (2013). Data Before the Fact. In L. Gitelman (Ed.), 'Raw Data' is an Oxymoron (pp. 15-40). The MIT Press.
- Sacasas, L. M. (2021). The materiality of digital culture: It's not as invisible – or as disembodying – as our devices lead us to believe. *Comment*, 12 August. https://comment.org/the-materiality-of-digital-culture
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1), 1-12. https://doi.org/10.1177/2053951718820549
- Salehi, N., Irani, L. C., Bernstein, M. S., Alkhatib, A., Ogbe, E., Milland, K., & Clickhappier. (2015). We are dynamo: Overcoming stalling and friction in collective action for crowd workers.CHI '15, 18-23 April. http://dx.doi.org/10.1145/2702123.2702508
- Savitz, E. J. (2019). Google buys data-analytics firm Looker amid privacy concerns. Baron's, 6 June. www.barrons.com/articles/google-buys-data-analytics-firm-looker-privacy-concerns-51559837661
- Schoon, A., Mabweazara, H. M., Bosch, T., & Dugmore, H. (2020). Decolonising digital media research methods: Positioning African digital experiences as epistemic sites of knowledge production. *African Journalism Studies*, 41(4), 1-15. https://doi.org/10.1080/23743670.2020.1865645
- Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, 2019/01/29/). Fairness and Abstraction in Sociotechnical Systems. FAT* '19: *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59-68. https://doi.org/10.1145/3287560.3287598
- Shleifer, A. (2004). Does Competition Destroy Ethical Behavior? NBER Working Paper 10269. National Bureau of Economic Research. www.nber.org/papers/w10269
- Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research,* 7(1), 111-134. ww.jstor.org/stable/23010792
- Stiefel, L., Currie, M., Musiani, F., Sandoz, A., Silvast, A., & Williams, R. (2024). Preface: Governance by infrastructure. *First Monday*, 29(2), 1-4. https://firstmonday.org/ojs/index.php/fm/article/view/13559/11406
- Suber, P. (2012). Open Access. The MIT Press.
- Taylor, L., Mukiri-Smith, H., Petročnik, T., Savolainen, L., & Martin, A. (2022). (Re)making data markets: An exploration of the regulatory challenges. *Law, Innovation and Technology*, 14(2), 355-394. https://doi.org/10.1080/17579961.2022.2113671
- The Majority Report w/ Sam Seder. (2023). Data cartels and our data. www.youtube.com/watch?v=E-VWLIFUxIA
- Tucker, E. (2023). Our future inside the Fifth Column – Or, What chatbots are really for. Tech Policy Press, 14 June. www.techpolicy.press/our-future-inside-the-fifth-column-or-what-chatbots-are-really-for
- Turow, J., Lelkes, Y., Draper, N., & Waldman, A. E. (2023). *Americans Can't Consent to Companies' Use of Their Data: They Admit They Don't Understand It, Say They're Helpless to Control It, and Believe They're Harmed When Firms Use Their Data – Making What Companies Do Illegitimate*. https://ssrn.com/abstract=4391134 or http://dx.doi.org/10.2139/ssrn.4391134
- Tusikov, N. (2021). Internet Platforms Weaponizing Chokepoints. In D. W. Drezner, H. Farrell, & A, Newman (Eds), *The Uses and Abuses of Weaponized Interdependence* (pp. 133-148). Brookings Institution Press.
- Tuters, M., & Hagen, S. (2020). (((They))) rule: Memetic antagonism and nebulous othering on 4chan. *New Media & Society*, 22(12), 2218-2237. https://doi.org/10.1177/1461444819888746
- 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
- Van der Vlist, F., Helmond, A., & Ferrari, F. (2024). Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence. *Big Data & Society*, 11(1), 1-16. https://doi.org/10.1177/20539517241232630

- Veliz, C. (2021). *Privacy is Power: Why and How You Should Take Back Control of Your Data. Melville House*.
- Ven, K., Verelst, J., & Mannaert, H. (2008). Should you adopt open source software? *IEEE Software*, 25(3), 54-59. https://doi.org/10.1109/MS.2008.73
- Verhulst, S. G., & Schüür, F. (2023). Interwoven realms: Data governance as the bedrock for AI governance. *Medium*, Blog, 20 November. https://medium.com/data-policy/interwoven-realms-data-governance-as-the-bedrock-for-ai-governance-ffd56a6a4543
- Viljoen, S. (2021). A relational theory of data governance. Yale Law Journal, 131(2), 573-654. www.yalelawjournal.org/feature/a-relational-theory-of-data-governance
- Wang, W. W. (2023). China's digital transformation: Data-empowered state capitalism and social governmentality. *The African Journal of Information and Communication*, 2023(31), 1-13. https://doi.org/10.23962/ajic.i31.16296
- West, S. M. (2019). Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society*, 58(1), 20-41. https://doi.org/10.1177/0007650317718185
- Wikipedia. (2024). Sock puppet account. https://en.wikipedia.org/wiki/Sock_puppet_account
- Zammit, M., Voulgari, I., & Yannakakis, G. N. (2021). The Road to AI Literacy Education: From Pegagogical Needs to Tangible Game Design. University of Malta Paper for European Conference on Games Based Learning. www.academia.edu/78830274/The_Road_to_Ai_Literacy_Education_from_Pedagogical_Needs_to_Tangible_Game_Design
- Zhu, J. (2022). AI ethics with Chinese characteristics? Concerns and preferred solutions in Chinese academia. AI & Society, 39(3), 1261-1274. https://doi.org/10.1007/s00146-022-01578-w
- Zuboff, S. (2019). The Age of Surveillance Capitalism: *The Fight for a Human Future at the New Frontier of Power*. Profile Books.