AI and transparency

Ville Aula[1] and Tero Erkkilä[2]

[1] University of Helsinki & London School of Economics and Political Science
[2] University of Helsinki

Citation: Aula, V., & Erkkilä, T. (2024). "Chapter 13: AI and transparency". pp. 170-180. In Paul, R., Carmel, E., & Cobbe, J. (Eds.) Handbook on Public Policy and Artificial Intelligence. Cheltenham, UK: Edward Elgar Publishing. pp. 170-180. https://doi.org/10.4337/9781803922171.00020

Introduction

In this chapter we critically explore how the concept of transparency is used in the emerging scholarly and policy debates on AI systems, in particular, its political character. Calls for transparency are often motivated by an assumed link between transparency and accountability, but the relationship between the two is far from simple. Indeed, the concept of transparency has evolved over time, containing democratic and economic connotations, but is now gaining new technical interpretations in the context of AI systems.

This conceptual shift is important, because the emerging sociotechnical understandings of algorithmic transparency foregrounds specific framings of policy problems while deprioritizing others, therefore favouring the interests of some groups while others groups suffer negative consequences (cf. Bacchi 1999). How transparency is understood in the emerging AI policy debate therefore has far-reaching consequences to what solutions might be adopted and who they might benefit. Transparency of AI and algorithmic systems is therefore subject to a power struggle between actors that pursue different interests. As a political concept, algorithmic transparency hence carries the potential to be instrumentalized (Skinner 1989) for promoting technical and ethical solutions to AI instead of considering its broader democratic and economic aspects. The chapter identifies key ideas and arguments in these debates, paving way for further critical research into competing conceptualizations of algorithmic transparency in scholarly and policy debates.

We will begin our chapter by discussing the relevance of AI transparency. We will then contextualize the concept of transparency and its ideational history by discussing its development as a political ideal before discussing how it features in literature on AI systems and identifying emerging policies on AI transparency. We conclude that current debates on algorithmic transparency carry the promise of bringing societal and cultural aspects of AI to

debates on accountability, but fall short on establishing actual institutional arrangements through which civil society actors and key stakeholders could control the use of algorithmic and AI applications.

Why is AI transparency relevant?

There are several reasons why transparency of AI has emerged as a central theme in ethical and policy debates on AI. Here, we want to emphasize three interrelated drivers that have pushed transparency to the fore, with each driver prompting different responses: opacity of computational systems, platformization, and digital surveillance.

First, calls for AI transparency respond to algorithmic systems having become ubiquitous but remaining opaque to citizens. Data-driven algorithms and computational tools have been characterized as black boxes that can have a negative impact on citizens without their knowledge (Pasquale 2015). Calls for transparency have become stronger as more harms and injustices have been identified (O'Neill 2016; Noble 2018). Burrell (2016) distinguishes three varieties of AI opacity: (1) organizational opacity involved with secrecy of the organizations using AI, (2) technical opacity relating to the skills needed to make sense of new computational models, and (3) opacity of machine learning models and their operational environment. These varieties of opacity are relevant for public policy debate, because they indicate differences in what should be the object of interventions promoting transparency.

Second, transparency of AI has gained relevance also because of platformization: algorithmic systems are essential to digital businesses that collect, process, and monetize user data, but the ways this is done are often inscrutable for users (Srnicek 2017; van Dijck et al. 2018). Social media platforms became successful by establishing a digital infrastructure for social transactions, and corporations try to replicate this strategy of digital infrastructures in new fields (Plantin et al. 2018). Transparency of AI technologies therefore has a political economy element that cannot be reduced to the sociotechnical functioning of computational systems. Many policy interventions to promote AI transparency therefore target a handful of multinational corporations that control these platforms.

Third, calls for transparency respond to concerns of secretive government and corporate surveillance. The revelations by Edward Snowden in 2013 on the digital surveillance of the US National Security Agency initiated a critical debate on how online and digital data were used by governments (e.g. Lyon 2014). Domestic political surveillance and censorship by authoritarian governments increasingly relies on algorithmic solutions (e.g. King et al. 2012). However, the surveillance is no longer restricted to governments. When the business model of platform companies hinges on collecting as much data as possible, users become victims of

corporate surveillance that monetizes online transactional data (Couldry and Mejias 2019; Zuboff 2015). Greater transparency has been called to curb both government and corporate surveillance as an infringement of citizens' rights.

Opacity, platformization, and surveillance are empirical developments that anchor debates on AI transparency. Their importance is evident in transparency becoming a leading theme in AI ethics frameworks that attempt to formulate principles that would guide use of AI to socially desirable outcomes and mitigate risks and harms (Mittelstadt et al. 2016; Tsamados et al. 2022). Indeed, Jobin et al. (2019) found transparency to be the most popular principle to be included in ethical frameworks. However, including transparency in ethical guidelines or proposing transparency as a policy response to opacity does not yet explain what transparency is meant to *achieve.*

One of the key goals of promoting transparency is accountability (Cobbe and Singh, Chapter 7 in this volume). When demands of transparency link with demands of accountability, their aim is to regulate the use of AI systems. Some argue that transparency of AI of itself is a method of accountability; others argue that transparency is a preliminary step; and yet others propose that forms of accountability are contingent on what aspects of AI systems are made transparent (for an overview, see Wieringa 2020). Indeed, algorithmic transparency, even if fully achieved on the level of individual algorithms or data sets, might have only limited effect as an accountability mechanism (Ananny and Crawford 2018). Furthermore, mere knowledge of problems and harms might not be useful as an accountability mechanism if it does not fall within the remit of existing legal protections. Selective transparency can even eschew real accountability if it lulls users into false sense of security or obfuscates user perception of the AI systems. Because of such uncertainties, corporate interests play a major role in trying to shape public and policy debate on transparency.

In addition to accountability, transparency of AI systems can be used to promote trust in AI and the organizations using it (e.g. von Eschenbach 2021; Gillis, Laux and Mittelstadt, Chapter 14 in this volume). This goal has become prominent especially after the problems of opacity have eroded public trust, although trust would be needed for increased adoption of AI, which is a conundrum that the European Union AI Act tries to tackle with its goal of

trustworthy AI (Laux et al. 2024). Trust as the goal of transparency has widespread currency in scholarly debate, although it has been criticized for its ambiguity, difficulty in terms of operationalization, and the uncertain implications of transparency (Felzman et al. 2019; Laux et al. 2024). Trust can be misused, and people can trust actors that are fundamentally untrustworthy, making trustworthiness a complicated goal for transparency (Reinhardt 2023). Furthermore, promotion of trust in AI differs significantly from the goal of accountability, because it aims at popular acceptance of AI. This means that rules and regulations relating to

transparency are balanced against the goal of increased uptake, which creates tension with the goal of accountability.

Unpacking the trade-offs, stakeholder interests, and unintended consequences of concepts like transparency, trustworthiness, and accountability is a key priority for critical policy analysis on AI (Paul 2022). In the next section we show how the current framings of AI transparency have their roots in a longer trajectory of how transparency is understood as a political ideal, which is now being reinterpreted in the context of AI.

Shifts in transparency as a political ideal

While the word transparency is fairly recent in its current popular meaning (Hood 2006), the concepts of openness and publicity have long histories. To put the debate on transparency of AI systems into the broader political context needed for critical analysis, it is necessary to identify key ideas that precede current debates on AI transparency.

Historical accounts of institutional openness or "transparency" are characterized by concern with social conflicts between the respective roles and authorities of markets, (state) institutions, and citizen rights (Emirbayer and Sheller 1999; Habermas 1989; Schulz-Forberg and Stråth 2010). Transparency as a political concept has its roots in the Enlightenment, when it came to be associated with a form of rule that can and should be scrutinized by citizens (Hood 2006). The lineages of openness and state secrecy differ between countries and have been discussed in terms of path dependence and its critical junctures (Knudsen 2003). Yet, the 1766 Swedish law on public access to state information was for a long time an exception to the prevailing practice of bureaucratic secrecy (Konstari 1977; Knudsen 2003; compare Gestrich 1994). The Swedish act was linked to the new printing techniques, and granted the right to publish information relating to the state and government documents, a development crucial to the emergence of the "public sphere". Here it is important to notice that new notions of publicity were connected to the introduction of new communication technologies, a situation analogous to current debates on AI systems and digital platforms.

Nevertheless, the practical implications of Enlightenment transparency ideals are different depending on whether they refer to publicity of the public sphere, transparency of state bureaucracy, liberalism of an open market economy, or budding political openness of republicanism and democracy. This makes transparency and openness themselves subject to a power struggle between actors with different interests.

Since the mid-twentieth century, politicization of government, the computerization of public administration, and transnational communication of policy innovations has led to the spread of government transparency (Bennett 1997; Schudson 2015). In addition, the end of the Cold War and the opening of the global market economy have also greatly contributed to the rise of transparency in public administration (Best 2005; Rose-Ackerman 2005). Since the 1990s, the rise of the internet has created pressures for transparency. More recently, big data, social media, and algorithmic governance have again influenced states' information strategies and transparency of public administration. An important first link between computational technology and transparency was established in the emergence of Free Software/Open Source programming (Kelty 2008; Coleman 2012). Starting from the 1980s, software developers promoted the idea that programming source code should be publicly shared instead of being a private property. Initiatives like Open Science and Open Data grew out of the initial Open Source movement in the early 2000s, arguing that companies, researchers, and governments should share their data with the public. The motivation for these initiatives is the claim that openness leads to faster innovation and therefore to more benefits to society. In practice, however, attempts to promote open data have had ambiguous and even contradictory goals relating to democracy, the economy, and innovation (Janssen et al. 2012; Yu and Robinson 2011). Nevertheless, examples of the enduring appeal of these ideals is that one of the key AI development companies is called "OpenAI", and that AI researchers often collaborate in sharing some of their data sets and models.

The above ideals of transparency exercise ongoing influence on the struggle over AI policy and regulation. Citizens and civil society who face new harms from AI are locked in a power struggle with private corporations and developers who profit from AI systems, with each side using openness and transparency to make their case. Civil society can appeal to democratic ideals to demand regulation and transparency. Citizens can appeal to ideals of public scrutiny when resisting governmental decisions reached via AI. Private corporations using AI in their business can appeal to economic ideals to promote transparent market practices and resist regulation. Developers of AI can appeal to technological openness to accelerate the development of new products and services. Contrasting interests are clearly evident in this list of what transparency can be used to justify in debates on AI. This makes it necessary to apply a critical approach to the conceptualizations, political economy, and contingent applications of the (variable) ideals of AI transparency.

Transparency in literature on AI systems

A specialist literature on transparency of algorithmic and AI systems has emerged in the last ten years. In this section we further explore the political underpinnings of possible solutions to the problem of opacity in AI systems. Given the complex nature and vested interests in defining transparency in AI, it is no surprise that there are divergent views on *what* exactly

should be transparent in AI systems. Most importantly, the technical and definitional details of AI transparency are far from trivial due to the contrasting interests they might serve. As argued by Amoore (2020), AI systems consist of various dependencies between humans and machines that constantly modify their interaction and operation, making it impossible to say what would be the ultimate point of origin whose transparency would *alone* reveal why an algorithm gives a specific output. Rather there are multiple elements whose transparency can each reveal a partial perspective on the operations of the moving puzzle of AI systems (see also Burrell 2016; Ananny and Crawford 2018). Yet pinning down these elements is crucial for successful AI policy, making the issue an ongoing political struggle.

Transparency of machine learning models is one of the most intensely debated technical aspects of AI. In research literature this is discussed as a question of model interpretability and explainable AI. These two concepts are discussed in detail elsewhere in this *Handbook* (see Berry, Chapter 10 in this volume) and we will here focus on their link to the transparency of AI more broadly. The debate on model transparency is driven by the technical characteristics of some machine learning techniques being near-impossible to understand by humans. Researchers have proposed various techniques to deal with the problem, but there is no consensus on what constitutes interpretability and explainability, or what metrics should be used to measure the success of individual techniques (Lipton 2018; Carvalho et al. 2019). In addition to the technical layer of model opacity, problematic outcomes of AI systems might follow from the way a fully transparent algorithm interacts with specific data sets in specific operational environments (Ananny and Crawford 2018). Although researchers have developed techniques that enhance the transparency of AI systems, this does not guarantee that they are meaningful for citizens at large. Developers, users, regulators, and the general audience all have a different rationale for dealing with AI systems and need different things from transparency (Felzmann et al. 2019). Furthermore, solutions promoting transparency often lack a critical audience that could effectively scrutinize and challenge algorithmic decisions most ordinary citizens lack the necessary knowledge or resources to do so (Kemper and Kolkman 2019). As a result, researchers must be critical of whether framing transparency solely around explainability or interpretability of machine learning models serves larger transparency goals. The literature on algorithmic transparency further considers accessibility, which not only refers to the public availability of source code, but also external experts' ability to analyse the algorithm. Tested in an experimental scenario, explainability had a more positive effect on citizen trust in algorithmic governance than mere accessibility (Grimmelikhuijsen 2023).

In the event that an AI system remains opaque, some information on its effects can still be reached externally with "algorithmic audits". The goal of algorithmic audit is often not direct access into the AI systems themselves, but exploration of whether systematic analysis of their outcomes can reveal discrepancies, biases, or injustices in their operation (Sandvig et al. 2014). Such audits are promoted especially by civil society actors which can use them to reveal

biases and injustices, but algorithmic audits can also be used by government regulatory bodies to audit AI systems within government and in the private sector.

Research on AI systems often calls for transparency of the data used in AI systems. On one hand, there are demands for transparency into the data used to train AI systems (e.g. Hacker 2021; Bertino et al. 2019). Transparency of training data is meant to create opportunities to scrutinize the data sets and detect problems that would lead to systematic mistakes, inaccuracy, or bias in the models based on it. The assumption behind this is that detection of problems in the data will alleviate the problems and improve the AI systems. However, transparency of training data does not as such provide mechanisms of accountability. On the other hand, transparency has also been demanded to the operational data that guides individual decisions made by AI. This approach, however, faces obstacles because companies using AI systems are reluctant to share such data. Legitimate privacy concerns also limit the transparency of operational data beyond what can be handed to the information subjects. Furthermore, the business of collecting and monetizing personal data operates in a legal grey area where many politically suspect practices are not per se illegal, making transparency of data hardly a solution on its own (Crain 2018). The sheer volume and complexity of personal data used in AI systems, if made available to users, might in fact increase the *opacity* of the systems because it hides what matters for the algorithms in the seeming transparency of the data sets (Stohl et al. 2016). Transparency of data sets can therefore be a relevant avenue for developers of AI systems to improve their systems, but inadequate in providing a foundation for citizen redress or political action.

Calls for transparency in AI systems also extend to the corporate structures of the companies developing and using AI systems. The digital infrastructures and corporate dependencies underpinning the building, training, maintenance, and deployment of AI systems are highly complex and often hidden from the public. Not only are the structures geographically distributed across various jurisdictions, but also the complex vertical and horizontal dependencies across leading companies conceal key details of how AI systems are developed and who benefits from them (Ferrari 2023). Developers of AI systems can also obscure the human labour needed to train and maintain AI systems, making it unclear who is responsible for their development (Newlands 2021). Consequently, public authorities have difficulty in identifying AI systems as targets for policy intervention or regulation, and individual users have little understanding of what goes on behind the user interface. Lastly, public sector organizations using AI systems are often dependent on private technologies whose operational details their developers consider business secrets, making it difficult to determine responsibility for mistakes and harms caused by AI in the public sector.

The above discussion demonstrates that transparency of AI contains a variety of elements that provide a partial perspective to the working of the systems, with no guarantee that transparency alone will deliver the political goals that motivate the calls for openness. The

issue as to which of these aspects are inscribed into policies and regulations is therefore very salient. In the next section we discuss the emerging literature on transparency in AI policies.

Emerging policies of AI transparency

Although research literature on AI has developed new conceptual ideas of transparency, only some of them have started to make their way into practical policy. Governments across the world have taken different approaches to promotion and regulation of AI and it is not a given that transparency is treated as being important (Cath et al. 2018). The ways that governments address transparency in their AI policies are influenced by their political traditions and cultural values (Ahonen and Erkkilä 2020; on institutional filtering processes also see af Malmborg and Trondal, Chapter 5 in this volume). Furthermore, national policies can aim for transparency of different aspects of AI systems, making them a battleground for contrasting interests. Because the literature on AI transparency is still nascent and most policy interventions still in development, the coming years will provide a considerable opportunity for critical policy research.

First, there is scope to assess whether and how transparency is addressed in governmental policies promoting AI. Governments face contradictory pressures, both to promote the use of AI and to react to the challenges it poses. National AI strategies, for example, emphasize the opportunities of AI, discuss the need for ethical standards, and call for close public and private partnership (Radu 2021; Ulnicane et al. 2021). The notion of openness for the purpose of economic and technological progress can be more important in AI policies than the idea of transparency as political accountability. As discussed in this chapter and elsewhere in this volume, emphasis on ethics and a lack of government intervention can often undermine efforts to tackle the problems of AI systems.

In addition to the promotion of AI, some governments have crafted policies to regulate and deliberately increase transparency in AI systems. In the main, attempts to introduce algorithmic transparency have involved public descriptions of algorithm use in decision-making. In France, public bodies are expected to provide public descriptions of the algorithms they use in decision-making (Etalab 2021; Open Government Partnership 2021). There are also examples of public actors providing this information on their own initiative, for example the cities of Amsterdam and Helsinki (City of Amsterdam 2020; City of Helsinki 2020). The UK government has launched an Algorithmic Transparency Recording Standard (ATRS) that basically provides public organizations a format and mechanism for communicating their use of algorithmic tools in decision-making (UK Government 2023). This includes a centrally managed repository for reporting the functionality of the algorithm and the reason for its use.

In April 2023, the European Union launched the European Centre for Algorithmic Transparency (ECAT) to provide technical assistance and practical guidance (ECAT 2023). Residing under the European Commission's Joint Research Centre the ECAT aims to become an international hub for research and communicating best practices on algorithmic transparency (Bertuzzi 2023). The European Union General Data Protection Regulation (GDPR) was the first policy that directly tackled transparency of algorithmic systems (in addition to its primary goal of privacy), including goals regarding the right to explanation although this did not constitute a legal duty (Wachter et al. 2017). Nevertheless, the GDPR does include a duty of lawful, fair, and transparent processing of personal data, which also has implications for AI systems (Felzmann et al. 2019).

Overcoming the shortcomings of the GDPR is central to the proposed European Union AI Act, which takes a risk and harm-based approach to AI but might not contain explicit legal duties of transparency (Varošanec 2022). Further complications in relation to transparency arise from the fact that AI is often relevant in the context of digital platforms, which are subject to their own interventions by the European Union such as the Digital Services Act and Digital Markets Act (EUR-Lex 2022a; 2022b).

If the European Union has been active in regulating AI systems, the United States and China have opted for less interventionist policies. Academic research on how transparency is understood in these policies is, however, very limited. The dominant policy approaches in these countries have been the development and deployment of AI systems, not their regulation. Nevertheless, the different legal systems and regulatory cultures in Europe, the United States, and China mean that policymaking can also take different forms. The Chinese government has introduced several new policies on AI systems and digital platforms, but their practical implications for transparency are unclear. In the United States, AI policies have been developed in close collaboration with leading digital platforms and emphasize ethical frameworks. The development of such hybrid and networked forms of governance calls for critical policy analysis to examine how a transparency regime led by the private sector ultimately turns out and how its outcomes compare with other regimes. In the light of existing evidence, however, it is unlikely that private corporations that promote and profit from the proliferation of AI systems would voluntarily tackle the full complexity of problems relating to opacity, platformization, and surveillance (on AITs in labour regulation in these countries, see also Donoghue, Huanxin, Moore and Ernst, Chapter 26 in this volume).

Apart from regulating general use of AI systems, governments also have policies guiding transparency in their own use of AI. These discussions are a direct extension of classical debates and the rules of accountability and publicity in the governments context are often stronger in public administration than in private business. In the absence of access to private AI systems, governmental AI systems offer a unique window into how AI transparency policies and solutions work in practice. However, governments regularly use proprietary AI systems

and outsource services to private companies, which can again place practices beyond scholarly and public scrutiny. Again, analysis of only the technical layer of transparency is inadequate when AI systems themselves constitute a complex governance structure between public and private entities. The transparency of corporate ties, procurement practices, and interdependencies of public and private computational systems are therefore crucial aspects f scrutinizing AI systems in the public sector.

Critical analysis of AI transparency policies must consider who is participating in the debates on AI transparency and informing government policy. So far debates on AI transparency have been led by researchers close to the development and operationalization of AI systems, information law specialists, social media researchers, and theorists developing normative frameworks. Political scientists and public policy scholars have been largely absent from the transparency debate, although there are tensions between different notions of ransparency and uncertainty over the right policy instruments and governance structures.

Conclusion

As is clear from this chapter, literature on AI transparency often balances between improving transparency of AI systems and critiquing the opacity of AI systems. The transparency of AI remains caught between the sociotechnical complexity of algorithmic systems and the political ideals that make transparency desirable in the first place. Algorithmic transparency marks an ideational shift to the conceptual history of government transparency. While transparency has previously carried both democratic and market connotations, the debates on opacity of computational systems, platformization, and surveillance have added new technical elements to its conceptualization while highlighting the ethical issues of AI. This ideational shift is also apparent in the standing scholarship on the transparency of algorithmic and AI systems.

We identify a tension between the new sociotechnical conceptualization of transparency, adopted by scholars and practitioners alike, and the previous perceptions of transparency that perceived it primarily as a concept of democracy and markets. While the technological aspects of algorithmic transparency as well as the perspectives of AI bias, fairness, and equality are very important, these nevertheless frequently make the individual problems visible without providing a tangible mechanism of accountability (cf. Mulgan 2000). Scholars have argued that it is difficult to find a suitable audience for algorithmic transparency (Ananny and Crawford 2018; Kemper and Kolkman 2019), but it is even more challenging to further establish actual mechanisms through which algorithmic transparency is embedded in a roader accountability system (cf. Erkkilä 2007).

There is an apparent need to consider the broader democratic and economic aspects of algorithmic governance and AI in different institutional contexts. Here civil servants and private companies are key actors, but the inclusion of civil society actors should also be a priority. The sociotechnical perspective on algorithmic transparency carries the promise of bringing the societal and cultural aspects of AI to the debate on accountability. But the key challenge remains to establish institutional arrangements through which such transparency would include civil society actors and key stakeholders in the accountability system with actual mechanisms for controlling the use of algorithmic and AI applications.

References

Ahonen, P., & Erkkilä, T. (2020). 'Transparency in algorithmic decision-making: Ideational tensions and conceptual shifts in Finland'. *Information Polity*, **25** (4), 419–432.

Amoore, L. (2020). *Cloud Ethics Algorithms and the Attributes of Ourselves and Others*. Durham, NC: Duke University Press.

Ananny, M., & Crawford, K. (2018). 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability'. *New Media & Society*, **20** (3), 973–989.

Bacchi, C. L. (1999). *Women, Policy and Politics: The Construction of Policy Problems*, London: Sage Publications.

Bennett, C. J. (1997). 'Understanding ripple effects: The cross-national adoption of policy instruments for bureaucratic accountability'. *Governance*, **10** (3), 213–233.

Bertino, E., Kundu, A., & Sura, Z. (2019). 'Data transparency with blockchain and AI ethics'. *Journal of Data and Information Quality (JDIQ)*, **11** (4), 1–8.

Bertuzzi, L. (2023). 'EU Launches Research Centre on Algorithmic Transparency'. www .Euractiv.Com. 19 April 2023. https:// www .euractiv .com/ section/ platforms/ news/ eu -launches -research -centre-on -algorithmic -transparency/ .

Best, J. (2005). *The Limits of Transparency: Ambiguity and the History of International Finance*, Ithaca, NY: Cornell University Press.

Burrell, J. (2016). 'How the machine "thinks": Understanding opacity in machine learning algorithms'. *Big Data & Society*, **3** (1).

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). 'Machine learning interpretability: A survey on methods and metrics'. *Electronics*, **8** (8), 832.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). 'Artificial intelligence and the "good society": The US, EU, and UK approach'. *Science and Engineering Ethics*, **24** (2), 505–528.

City of Amsterdam. (2020). 'Amsterdam Algoritmeregister'. https:// algoritmeregister .amsterdam .nl/ en/ai -register/ .

City of Helsinki. (2020). ‘City of Helsinki AI Register’. https:// ai .hel .fi/ en/ ai -register/ .

Coleman, E. G. (2012). ‘Coding freedom’. In *Coding Freedom*, Princeton, NJ: Princeton University Press.

Couldry, N., & Mejias, U. A. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*, Stanford, CA: Stanford University Press.

Crain, M. (2018). ‘The limits of transparency: Data brokers and commodification’. *New Media & Society*, **20** (1), 88–104. ECAT. (2023). ‘European Centre for Algorithmic Transparency’. 9 October 2023. https:// algorithmic -transparency .ec .europa .eu/ index en.

Emirbayer, M., & Sheller, M. (1999). ‘Publics in history’. *Theory and Society*, **28** (1), 145–97.

Erkkilä, T. (2007). ‘Governance and accountability – A shift in conceptualisation’. *Public Administration Quarterly*, **31** (1), 1–38.

Etalab. (2021). ‘Fiche pratique: l’inventaire des principaux traitements algorithmiques’. 11 February. https:// guides .etalab .gouv .fr/ algorithmes/ inventaire/ #dans -quels -cas -une -administration -doit -elle - realiser -un -inventaire -de -ses -algorithmes.

EUR-Lex. (2022a). *Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on Contestable and Fair Markets in the Digital Sector and Amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA Relevance)*. *OJ L*. Vol. 265.http:// data .europa .eu/ eli/ reg/ 2022/ 1925/ oj/ eng.

EUR-Lex. (2022b). *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and Amending Directive 2000/31/EC (Digital Services Act) (Text with EEA Relevance)*. *OJ L*. Vol. 277. http:// data .europa .eu/ eli/ reg/ 2022/ 2065/ oj/eng.

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). ‘Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns’. *Big Data & Society*, **6** (1).

Ferrari, F. (2023). ‘Neural production networks: AI’s infrastructural geographies’. *Environment and Planning F*, **2** (4), 459–476. https:// doi .org/ 10 .1177/ 26349825231193226.

Gestrich, A. (1994). *Absolutismus Und Öffentlichkeit. Politische Kommunikation in Deutschland Zu Beginn Des 18. Jahrhunderts*, Göttingen: Vandenhoeck & Ruprecht.

Grimmelikhuijsen, S. (2023). ‘Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making’. *Public Administration Review* **83** (2), 241–262. https:// doi .org/ 10 .1111/ puar .13483.

Habermas, J. (1989). *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*, London: Polity Press.

Hacker, P. (2021). ‘A legal framework for AI training data – From first principles to the Artificial Intelligence Act’. *Law, Innovation and Technology*, **13** (2), 257–301.

Hood, C. (2006). 'Transparency in historical perspective', in Hood, C., & Heald, D. (eds), *Transparency: The Key to Better Governance?*, Proceedings of the British Academy, Oxford: Oxford University Press, pp. 1–24.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). 'Benefits, adoption barriers and myths of open data and open government'. *Information Systems Management*, **29** (4), 258–268.

Jobin, A., Ienca, M. & Vayena, E. (2019). 'The global landscape of AI ethics guidelines'. *Nature Machine Intelligence*, **1**, 389–399. https:// doi .org/ 10 .1038/ s42256 -019 -0088 -2.

Kelty, C. M. (2008). *Two Bits: The Cultural Significance of Free Software*, Durham, NC: Duke

University Press. Kemper, J., & Kolkman, D. (2019). 'Transparent to whom? No algorithmic accountability without a critical audience'. *Information, Communication & Society*, **22** (14), 2081–2096.

King, G., Pan, J., & Roberts, M. E. (2013). 'How censorship in China allows government criticism but silences collective expression'. *American Political Science Review*, **107** (2), 326–343.

Knudsen, T. (2003). *Offentlighed i Det Offentlige. Om Historiens Magt*, Aarhus: Aarhus Universitetsforlag.

Laux, J., Wachter, S., & Mittelstadt, B. (2024). 'Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk'. *Regulation & Governance*, **18** (1), 3–32. https:// doi .org/ 10 .1111/ rego .12512.

Lipton, Z. C. (2018). 'The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery'. *Queue*, **16** (3), 31–57.

Lyon, D. (2014). 'Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big Data & Society*, **1** (2), 2053951714541861.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). 'The ethics of algorithms: Mapping the debate'. *Big Data & Society*, **3** (2).

Mulgan, R. (2000). '"Accountability": An ever-expanding concept?'. *Public Administration*, **78** (3), 555–573.

Newlands, G. (2021). 'Lifting the curtain: Strategic visibility of human labour in AI-as-a-Service'. *Big Data & Society*, **8** (1), 20539517211016026.

Noble, S. U. (2018). *Algorithms of Oppression*, New York: New York University Press.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Penguin Books.

Open Government Partnership. (2021). 'Building public algorithm registers: Lessons learned from the French approach'. Open Government Partnership. 12 May 2021. https://www.opengovpartnership.org/stories/building-public-algorithm-registers-lessons-learned-from -the-french-approach/

Pasquale, F. (2015). *The Black Box Society*: *The Secret Algorithms that Control Money and Information*, Cambridge, MA: Harvard University Press.

Paul, R. (2022). 'Can critical policy studies outsmart AI? Research agenda on artificial intelligence technologies and public policy'. *Critical Policy Studies*, **16** (4), 497–509.

Plantin, J. C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). 'Infrastructure studies meet platform studies in the age of Google and Facebook'. *New Media & Society*, **20** (1), 293–310.

Radu, R. (2021). 'Steering the governance of artificial intelligence: National strategies in perspective'. *Policy and Society*, **40** (2), 178–193.

Reinhardt, K. (2023). 'Trust and trustworthiness in AI ethics'. *AI Ethics*, **3**, 735–744. https:// doi .org/ 10.1007/ s43681 -022 -00200 -5.

Rose-Ackerman, S. (2005). *From Elections to Democracy: Building Accountable Government in Hungary and Poland*, New York: Cambridge University Press.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). 'Auditing algorithms: Research methods for detecting discrimination on internet platforms'. *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, **22**, 4349–4357.

Schudson, M. (2015). *The Rise of the Right to Know: Politics and the Culture of Transparency, 1945–1975*, Cambridge, MA: Belknap Press: An Imprint of Harvard University Press.

Schulz-Forberg, H., & Stråth, B. (2010). 'Soft and strong European public spheres', in R. Frank, H. Kaelble, M. Lévy, & L. Passerini (eds), *Building a European Public Sphere: From the 1950s to the Present*, Brussels: PIE-Peter Lang, pp. 55–76.

Skinner, Q. (1989). 'Language and political change', in T. Ball, J. Farr, & R. L. Hanson (eds), *Political Innovation and Conceptual Change*, Cambridge: Cambridge University Press, pp. 6–23.

Srnicek, N. (2017). *Platform Capitalism*, Polity. Cambridge.

Stohl, C., Stohl, M., & Leonardi, P. M. (2016). Managing opacity: Information visibility and the paradox of transparency in the digital age. *International Journal of Communication*, **15** (10), 123–137.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2022). 'The ethics of algorithms: Key problems and solutions'. *AI & Society*, **37** (1), 215–230.

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W. G. (2021). 'Framing governance for a contested emerging technology: Insights from AI policy'. *Policy and Society*, **40** (2), 158–177.

UK Government. (2023). 'Algorithmic Transparency Recording Standard – Guidance for Public Sector Bodies'. GOV.UK. 5 January. https://www.gov.uk/ government/publications/ guidance -for-organisations-using-the-algorithmic-transparency-recording-standard/algorithmic-transparency-recording-standard-guidance-for-public-sector-bodies.

Van Dijck, J., Poell, T., & De Waal, M. (2018). *The Platform Society: Public Values in a Connective World*, Oxford: Oxford University Press.

Varošanec, I. (2022). 'On the path to the future: Mapping the notion of transparency in the EU regulatory framework for AI'. *International Review of Law, Computers & Technology*, **36** (2), 95–117, DOI:10.1080/13600869.2022.2060471.

von Eschenbach, W. J. (2021). 'Transparency and the black box problem: Why we do not trust AI'. *Philosophy & Technology*, **34** (4), 1607–1622.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). 'Why a right to explanation of automated decision-making does not exist in the general data protection regulation'. *International Data Privacy Law*, **7** (2), 76–99.

Wieringa, M. (2020, January). 'What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability', in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 1–18. https:// doi .org/ 10 .1145/ 3351095 .3372833.

Yu, H., & Robinson, D. G. (2011). 'The new ambiguity of open government'. *UCLA Law Review Discourse*, **59**, 178–208.

Zuboff, S. (2015). 'Big other: Surveillance capitalism and the prospects of an information civilization'. *Journal of Information Technology*, **30** (1), 75–89.