Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges, An Introduction

Corinne Cath[1,2]

[1] Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford, OX1 3JS, United Kingdom.

[2] The Alan Turing Institute, 96 Euston Road, London NW1 2DB, United Kingdom.

Correspondence author email: ccath@turing.ac.uk

Abstract

This paper is the introduction to the special issue entitled: "Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges". Artificial Intelligence (AI) increasingly permeates every aspect of our society, from the critical, like urban infrastructure, law enforcement, banking, healthcare, and humanitarian aid, to the mundane like dating. AI, including embodied AI in robotics and techniques like machine learning, can improve economic, social welfare, and the exercise of human rights. Due to the proliferation of AI in high-risk areas, the pressure is mounting to design and govern AI to be accountable, fair and transparent. How can this be achieved and through which frameworks? This is one of the central questions addressed in this special issue, in which eight authors present an in-depth analysis of the ethical, legal-regulatory, and technical challenges posed by developing governance regimes for AI systems. It also gives a brief overview of recent developments in AI governance, how much of the agenda for defining AI regulation, ethical frameworks, and technical approaches is set, as well as providing some concrete suggestions to further the debate on AI governance.

Keywords: Artificial Intelligence (AI), Law, Ethics, Technology, Governance, Culture

Introduction

Artificial Intelligence (AI) increasingly permeates every aspect of our society, from the critical, like healthcare, and humanitarian aid, to the mundane like dating. AI, including embodied AI in robotics and techniques like machine learning, can enhance economic, social welfare, and the exercise of human rights. The various sectors mentioned can benefit from these new technologies. At the same time, AI may be misused or behave in unpredicted and potentially harmful ways. Questions on the role of the law, ethics, and technology in governing AI systems are thus more relevant than ever before. Or, as Floridi (1) argues in this issue: 'because the digital revolution transforms our views about values and priorities, good behaviour, and what sort of innovation is not only sustainable but socially preferable – and governing all this has now become the fundamental issue' (p.2).

AI systems, most of which apply learning techniques from statistics to find patterns in large sets of data and make predictions based on those patterns, are used in a wide variety of applications. Due to the proliferation of AI in high-risk areas, pressure is mounting to design and govern AI to be accountable, fair and transparent. How can this be achieved and through which frameworks? This is one of the central questions addressed by the various authors in this special issue, who present an in-depth analysis of the ethical, legal-regulatory, and technical challenges posed by developing governance regimes for AI systems.

Societies are increasingly delegating complex, risk-intensive processes to AI systems, such as granting parole, diagnosing patients, and managing financial transactions. This raises new challenges, for example, around liability regarding automated vehicles, the limits of current legal frameworks in dealing with 'big data's disparate impact' (2) or preventing algorithmic harms (3), social justice issues related to automating law enforcement or social welfare (4), or online media consumption (5). Given AI's broad impact, these pressing questions can only be successfully addressed from a multidisciplinary perspective.

This theme issue collects eight original articles, written by internationally leading experts in the fields of AI, computer science, data science, engineering, ethics, law, policy, robotics, and social sciences. The articles are revised versions of papers presented at three workshops organised in 2017 and 2018 by Corinne Cath, Sandra Wachter, Brent Mittelstadt, and Luciano Floridi (the editors) at the Oxford Internet Institute and the Alan Turing Institute. The workshops were titled: "Ethical Auditing for Accountable Automated Decision-making"; "Ethics & AI: Responsibility & Governance"; and "Explainable and Accountable Algorithms". This special issue will propose new ideas on how to develop and support the ethical, legal, and technical governance of AI. It is focused on the investigation of three specific areas of research:

(1) Ethical governance: focusing on the most pertinent ethical issues raised by AI, covering issues such as fairness, transparency, and privacy (and how to respond when the use of AI can lead to large-scale discrimination), the allocation of services and goods (the use of AI by industry, government, and companies), and economic displacement (the ethical response to the disappearance of jobs due to AI based automation).

(2) Explainability and interpretability: these two concepts are seen as possible mechanisms to increase algorithmic fairness, transparency, and accountability. For example, the idea of a "right to explanation" of algorithmic decisions is debated in Europe. This right would entitle individuals to obtain an explanation if an algorithm decides about them (e.g. refusal of loan application). However, this right is not yet guaranteed. Further, it remains open how we would construe the "ideal algorithmic explanation" and how these explanations can be embedded in AI systems.

(3) Ethical auditing: for inscrutable and highly complex algorithmic systems, accountability mechanisms cannot solely rely on interpretability. Auditing mechanisms are proposed as possible solutions that examine the inputs and outputs of algorithms for bias and harms, rather than unpacking how the system functions.

A growing body of literature covers questions of AI and ethical frameworks (1,6–10), laws (3,11–14) to govern the impact of AI and robotics (15), technical approaches like algorithmic impact assessments (16–18), and building trustworthiness through system validation (19). These three guiding forces in AI governance (law, ethics, technology) can be complementary (1). However, the debate on when which approach (or combination of approaches) is most relevant is unresolved, as Nemitz and Pagallo expertly highlight in this issue (13,17).

Across the globe, industry representatives, governments, academics, and civil society are debating where legal-regulatory frameworks are needed and when, if ever, ethical or technical approaches suffice. Even if those questions are answered, the question of the extent to which our existing ethical and regulatory frameworks sufficiently cover the impact of these technologies remains. Pagallo in this issue, for instance, highlights this conundrum by analysing the debate on the legal status of embodied AI (robots) in the EU (14). Veale, Binns, and Edwards argue here that European data protection provides robust principles but that

‘many socio-technical challenges presented by machine learning and algorithmic systems more broadly are not wholly dealt with using the provisions in regulations such as the General Data Protection Regulation (GDPR), which are the result of a slow evolution in definitions and concerns’ (p.17) (3). Winfield and Jirotka specifically consider the role of technical standards in the ethical and agile governance of robotics and AI systems.

Academia is also debating its own approach to AI governance. In a recent article on ‘troubling trends in machine learning scholarship’ Lipton and Steinhardt (24), for example, warned against technical solutionism through the misuse of concepts like ‘fairness’ and ‘discrimination’. They argue that borrowing these complicated social concepts to talk about ‘simple statistics’ is dangerous because it is ‘confusing researchers who become oblivious to the difference, and policymakers who become misinformed about the ease of incorporating ethical desiderata into machine learning’ (p. 5). Various academics expertly questioned the imaginaries underlying data-driven technologies like AI (20) in current debates and highlighted the risks of the use of AI systems (21–23). More work needs to be done to apply these critical lenses to the ethical, legal, and technical solutions proposed for AI governance.

The articles in this special issue reflect the nuanced and advanced state of the debate. At the same time, the authors also show that some of the legal governance solutions proposed are too limited in scope. As well as indicating that particular ethical solutions suffer from conceptual ambiguity and lack of enforcement mechanisms. Likewise, some technical approaches run the risk of narrowing down complicated social concepts, like fairness, beyond recognition or turning transparency into a box-ticking exercise. Hence, in addition to suggesting further ethical, legal, and technical refinements, the articles in this special issue also critically assess the status-quo of AI governance. In doing so, the authors highlight the importance of considering who is driving AI governance and what these individuals and organisations stand to gain. Because as Harambam, Helberger and Van Hoboken (5) state in this issue: ‘Technology is, after all, never an unstoppable or uncontrollable force of nature, but always the product of our making, including the course it may take. Even with AI’ (p.1).

There are clearly outstanding questions regarding what good AI governance should look like (2,25–27). These questions are currently debated by political institutions across the globe,

including the UK[1], South Korean[2], the Indian[3], and the Mexican government[4], as well as the European Commission.[5] Through the articles in this special issue we hope to contribute to shaping these debates. To situate the various articles, a brief overview of recent developments in AI governance and how agendas for defining AI regulation, ethical frameworks, and technical approaches are set, will be given.

Setting the Agenda for AI Governance

Academics and regulators alike are scrambling to keep up with the number of articles, principles, regulatory measures, and technical standards produced on AI governance. In the first six months of 2018 alone, at least a dozen countries put forward new AI strategies[6], several pledging up to 1,8 billion[7] in government funding. Industry, meanwhile, is developing its own AI principles[8] or starting multistakeholder initiatives to develop best-practices. They are also involved in developing regulation for AI, whether through direct participation or lobbying efforts. These industry efforts are laudable, but it is important to position them in light of three important questions. First, who sets the agenda for AI governance? Second, what cultural logic is instantiated by that agenda and, third, who benefits from it? Answering these questions is important because it highlights the risks of letting industry drive the agenda and blind spots in current research efforts.

Excellent work exists on the problematic developments in machine learning research regarding the conflation of complicated social concepts with simple statistics (24,28). Similarly, various authors highlight how unchecked use of 'black box' systems in finance (29), education and criminal justice (30), search engines (31) or social welfare (4) can have

[1] House of Commons. 2017. Algorithms in decision-making. House of Commons Science and Technology Committee. https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf

[2] Korea Times. 2018. Korea to nurture artificial intelligence, robot as new cash cow. https://www.koreatimes.co.kr/www/news/biz/2016/10/123_217161.html

[3] Times of India. 2018. NITI Aayog releases strategy on artificial intelligence, identifies 5 focus areas. NITI Aayog releases strategy on artificial intelligence, identifies 5 focus areas http://timesofindia.indiatimes.com/articleshow/64452469.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst

[4] Mexican Government. 2018. Estrategia de Inteligencia Artificial MX 2018. https://datos.gob.mx/blog/estrategia-de-inteligencia-artificial-mx-2018?category=noticias&tag=nula

[5] European Commission. 2018. Artificial intelligence: Commission outlines a European approach to boost investment and set ethical guidelines. http://europa.eu/rapid/press-release_IP-18-3362_en.htm

[6] Dutton, T. (2018). Politics of AI, an overview of national AI strategies. https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

[7] Reuters. 2018. France to spend 1.8 billion on Ai to compete with U.S., China. https://www.reuters.com/article/us-france-tech/france-to-spend-1-8-billion-on-ai-to-compete-with-u-s-china-idUSKBN1H51XP

[8] AI at Google: Our Principles. https://www.blog.google/technology/ai/ai-principles/

detrimental effects. Beer (32) aims to focus the debate on the 'social power of algorithms'. He argues that the cultural notion of algorithm serves 'as part of the discursive reinforcement of particular norms, approaches and modes of reasoning' (p. 11). As mentioned, it is not just how AI systems work, but also how we understand and imagine them (20) that fundamentally shapes AI governance. The next paragraphs will highlight some concerns and invite closer scrutiny of the cultural logic set forward by having industry actively shape the debate.

Many of the industry leaders in the field of AI are incorporated in the U.S. An obvious concern is the extent to which AI systems mirror societies in the image of U.S. culture and to the predilections of American tech behemoths. AI programming does not necessarily require massive resources. Much of its value comes from the data that is held. As a result, most of the technical innovation is led by a handful of American companies.[9] As these companies are at the forefront of various regulatory initiatives[10], it is essential to ensure this particular concern is not exacerbated. An American, corporate needs-driven agenda is not naturally going to be a good fit for the rest of the world. For instance, the EU has very different privacy regulations than the US. But this is not the only concern.

AI systems are often presented as being 'black boxes' (29) that are very complex and difficult to explain (22). Kroll in this issue shows that these arguments obfuscate that algorithms are fundamentally understandable (19). He argues that 'rather than discounting systems which cause bad outcomes as fundamentally inscrutable and therefore uncontrollable, we should simply label the application of inadequate technology what it is: malpractice, committed by a system's controller' (p.5). Yet, the cultural logic of the 'complicated inscrutable' technology is often used to justify the close involvement of the AI industry in policy making and regulation.[11] Generally, the industry players involved in these policy processes represent the same select group that is leading the business of online marketing and data collection. This is not a coincidence. Companies like Google, Facebook, and Amazon are able to gather large quantities of data, which can be used to propel new AI-based services. The

[9] We recognize that there are various major technical players in China and other Asian countries that play a significant role in furthering technological developments in the field of machine learning. However, these companies play a less prominent role in global policy development regarding AI governance than American companies.

[10] See for instance: Independent. 2014. Lobbyists demands were copied into law by meps. https://www.independent.co.uk/news/world/europe/lobbyists-demands-were-copied-into-law-by-meps-8493933.html, and the corporate sponsorship of the annual computers privacy and data protection (CPDP) conference in Belgium http://www.cpdpconferences.org/sponsorship.html, or the main founders of the Partnership on AI initiative https://www.partnershiponai.org/

[11] ZDNet 2018. UK enlists Deepmind's Demis Hassabis to advise its new Government Office for AI. https://www.zdnet.com/article/uk-enlists-deepminds-demis-hassabis-to-advise-its-new-government-office-for-ai/

'turn to AI' thus both further consolidates big companies' market position and provides legitimacy to their inclusion in regulatory processes.

A related concern is the influence companies exert over AI regulation. In some instances, they act as semi-co-regulators. For example, after the Cambridge Analytica scandal, Facebook's CEO testified before a joint-hearing of the US Senate Commerce and Judiciary Committee about his company's role in the data breach. During the hearing, he was explicitly asked[12] by multiple Senators to provide examples of what regulation for his company should look like. Likewise, the European Commission recently appointed a High-Level Expert Group (HLEG) on AI.[13] The group is mandated to work with the Commission on the implementation of a European AI strategy. The group's 52 members come from various backgrounds and, even though not all affiliations are apparent, it appears almost half of the members are from industry; 17 are from academia, only four are from civil society. Marda, in this issue, highlights the importance of ensuring civil society – often closest to those affected by AI systems – has an equal seat at the table when developing AI governance regimes. She shows that the current debate in India is heavily focused on governmental and industry concerns and goals of innovation and economic growth, at the expense of social and ethical questions (27).

Nemitz, likewise, focuses on how a limited number of corporations wield a lot of power in the field of AI. He states in this issue: 'The critical inquiry into the relationship of the new technologies like AI with human rights, democracy and the rule of law must therefore start from a holistic look on the reality of technology and business models as they exist today, including the accumulation of technological, economic and political power in the hands of the "frightful five", which are at the core of the development and systems integration of AI into commercially viable services.' Industry's influence is also visible in the creation of various large-scale global initiatives on AI and ethics. There are clear advantages to having open norm-setting venues that aim to address AI governance by developing technical standards, ethical principles, and professional codes of conducts. However, the solutions presented could do more to go beyond current voluntary ethical frameworks or narrowly defined technical interpretations of fairness, accountability, and transparency. The various articles in this edition clearly indicate why it is vital to further address questions of hard-regulation or the internet's

[12] Tech Crunch. 2018. Zuckerberg testified at congressional hearings. https://techcrunch.com/story/zuckerberg-testifies-at-congressional-hearings/

[13] European Commission 2018. High-Level Group on Artificial Intelligence. https://ec.europa.eu/digital-single-market/en/high-level-group-artificial-intelligence

business model of advertising and attention. If we are serious about AI governance, then these issues must be holistically contended with.

Concluding Remarks

The argument presented in this article should not be read as a dismissal of the work done by industry or the relevance of current ethical, technical solutions, and regulatory AI governance frameworks. Rather, much can be learnt from this ongoing work but only if we carefully assess its aims, impact, and process. It is crucial to remain critical of the underlying aims of AI governance solutions as well as the (unforeseen) collateral cultural impacts, especially in terms of legitimising private-sector led norm development around ethics, standards, and regulation. Likewise, we must remain cognizant of the concerns not, or only partially, covered by phrases like fairness, accountability, and transparency. In focusing on these issues what is not discussed? Are we assuming that issues around AI and equity, social justice, or human rights are automatically caught by these popular acronyms? Or are these concerns out of scope for the organisations pushing the agenda? Asking these hard questions matters because these concepts are increasingly making their way into regulatory initiatives[14] across the globe.

The authors in this special issue expertly engage with these various hard questions. From the articles, it becomes clear that the authors are unsatisfied with the current state of AI governance. Nemitz, for instance, argues in favour of fostering a new culture of technology and business development stooled on the rule of law, human rights, and democratic principles (17). Pagallo highlights the importance of pragmatism and testing new forms of accountability and liability through methods of legal experimentation (14). Veale, Binns, and Edwards explore how machine learning models could be considered personal data under European data protection law and argue that 'enabling users to deploy local personalisation tools might balance power relations in relation to large firms hoarding personal data' (p.5) (3). Winfield and Jirotka argue **that creating strong ethical principles is only the first step and that** more should be done to assure implementation and accountability. **Because** the real test for good governance **of AI systems** comes when the rubber hits the road, or rather, the robot.

Harambam, Helberger, and Van Hoboken explore the notion of 'voice', both as a way of allowing individuals to exert more control over the algorithms in the news industry and to mitigate the pitfalls of attempts at achieving algorithmic transparency (5). The editors argued

[14] French National AI Strategy. 2018. https://www.aiforhumanity.fr/en/

here, and in other pieces (26), that it is important to ensure that there is equitable stakeholder representation when regulating AI. Furthermore, there is a need for more non-U.S. led initiatives like the Europe-based AI4People[15] and the Council on Europe's Expert Committee on AI and Human Rights.[16] Even though it is important to have more Europe-led initiatives, we must also incorporate concerns from the Global South. Marda's article about India highlights why these voices are especially relevant (27). Similarly, it is essential to go beyond the fairness, accountability, and transparency rhetoric to formulate what additional fundamental values should be included. Nemitz, Floridi, and Marda, for example, argue for the inclusion of human rights' principles (1,17,27).

Overall, the critical perspectives offered in this special issue highlight the nuances of the debate on AI, ethics, technology, and the law and pave the road for a broader, more inclusive, AI governance agenda. Or as Kroll reminds us: 'In general, opacity in sociotechnical systems results from power dynamics between actors that exist independent of the technical tools in use. No artefact [sic] is properly comprehended without reference to its human context, and software systems are no different' (p. 11) (19).

The editors would like to thank the authors for their thoughtful engagement with the topics of this special issue. Their contributions are exemplary of the kind of multi-disciplinary research needed. Here, an attempt was made tried to highlight the various topics covered by the authors, but the short summaries included do not do justice to the rich and invigorating arguments made in the individual articles. The articles both reflect the three central themes of this special issue: ethical governance, explainability and interpretability, and ethical auditing as well as critically assessing the current state of AI governance.

We would like to thank the Oxford Internet Institute (OII), the Alan Turing Institute (ATI), and in particular the ATI's Data Ethics Group (DEG) for supporting the workshops that led to this Special Issue. We would also like to express our gratitude to the PETRAS Internet of Things research hub[17] for their support. Throughout this special issue the reader is invited to, as Floridi argues, resist the distracting narrative that 'digital innovation leads, and everything

[15] http://www.eismd.eu/ai4people/ Disclosure, one of the editors of this special issue is part of the AI4people initiative.

[16] https://www.coe.int/en/web/freedom-expression/msi-aut, Disclosure, one of the editors of this special issue is part of the Council of Europe expert committee.

[17] Cath's and Floridi's contributions to the editing of this theme issue have been funded as part of the Privacy and Trust Stream - Social lead of the PETRAS Internet of Things research hub. PETRAS is funded by the Engineering and Physical Sciences Research Council (EPSRC), grant agreement no. EP/N023013/1. All authors declare no conflicts of interest.

else lags behind, or follows at best: business models, working conditions, standards of living, legislation, social norms, habits, expectations and even hope' (p. 2) (1).

References

1. Floridi L. Soft Ethics and the Governance of the Digital. Cath C, Wachter S, Mittelstadt B, Floridi L, editors. Philosophical Transactions of the Royal Society A. Forthcoming;(Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges).

2. Barocas S, Selbst AD. Big Data's Disparate Impact. 104 CALIF L REV 671 [Internet]. 2016 [cited 2018 May 9]; Available from: http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf

3. Veale M, Binns R, Edwards L. Algorithms that Remember: Model Inversion Attacks and Data Protection Law. Cath C, Wachter S, Mittelstadt B, Floridi L, editors. Philosophical Transactions of the Royal Society A. Forthcoming;(Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges).

4. Eubanks V. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York, NY: St. Martin's Press; 2018.

5. Harambam J, Helberger N, Van Hoboken J. Democratizing Algorithmic News Recommenders: how to materialize voice in a technologically saturated media ecosystem. Cath C, Wachter S, Mittelstadt B, Floridi L, editors. Philosophical Transactions of the Royal Society A. Forthcoming;(Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges).

6. Floridi L, Taddeo M. What is data ethics? Philos Trans A Math Phys Eng Sci [Internet]. 2016 [cited 2018 May 8];374(2083). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5124072/

7. Ananny M. Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. Science, Technology, & Human Values. 2016;41(1):93–117.

8. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The Ethics of Algorithms: Mapping the Debate. Big Data & Society [Internet]. 2016; Available from: http://philpapers.org/rec/MITTEO-12

9. Winfield A, Jirotka M. Ethical Governance is Essential to Building Trust in Robotic and AI Systems. Cath C, Wachter S, Mittelstadt B, Floridi L, editors. Philosophical Transactions of the Royal Society A. Forthcoming;

10. Taddeo M, Floridi L. How can AI be a force for good: an ethical framework will help harness the potential of AI while keeping humans in control. Science Perspectives. 2018;361(6404).

11. Wachter S, Mittelstadt B, Floridi L. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law. 2017;7(2):76–99.

12. Edwards L, Veale M. Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'? IEEE Security and Privacy Magazine [Internet]. 2018 [cited 2018 Jul 17]; Available from: http://discovery.ucl.ac.uk/10042153/1/SSRN-id3052831%281%29.pdf

13. Pagallo U. Apples, Oranges, Robots: Four Misunderstandings in Today's Debate on the Legal Status of AI Systems. Cath C, Wachter S, Mittelstadt B, Floridi L, editors. Philosophical Transactions of the Royal Society A. Forthcoming;Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges.

14. Mendoza I, Bygrave LA. The Right Not to Be Subject to Automated Decisions Based on Profiling. University of Oslo Faculty of Law Research Paper [Internet]. 2017 [cited 2018 Aug 27];2017–20. Available from: https://papers.ssrn.com/abstract=2964855

15. Wachter S, Mittelstadt B, Floridi L. Transparent, explainable, and accountable AI for robotics. Science Robotics. 2017;2(6):eaan6080.

16. Selbst AD. Disparate Impact in Big Data Policing. Georgia Law Review [Internet]. 2017 [cited 2018 May 24];52(109). Available from: https://www.georgialawreview.org/article/3373-disparate-impact-in-big-data-policing

17. Nemitz P. Constitutional Democracy and Technology in the age of Artificial Intelligence. Cath C, Wachter S, Mittelstadt B, Floridi L, editors. Philosophical Transactions of the Royal Society A. Forthcoming;(Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges).

18. AI NOW. Algorithmic Impact Assessments: a practical framework for public agency accountability. New York, USA: AI NOW; 2018.

19. Kroll JA. The Fallacy of Inscrutability. Cath C, Wachter S, Mittelstadt B, Floridi L, editors. Philosophical Transactions of the Royal Society A. Forthcoming;(Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges).

20. Elish MC, boyd danah. Situating methods in the magic of Big Data and AI. Communication Monographs. 2017;85(1):57–80.

21. Citron DK. Technological Due Process [Internet]. Rochester, NY: Social Science Research Network; 2007 [cited 2018 May 9]. Report No.: ID 1012360. Available from: https://papers.ssrn.com/abstract=1012360.

22. Burrell J. How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society. 2016;3(1):2053951715622512.

23. Crawford K, Calo R. There is a blind spot in AI research. Nature News. 2016;538(7625):311.

24. Lipton ZCL, Steinhardt J. Troubling Trends in Machine Learning Scholarship. ICML 2018 debate papers [Internet]. 2018; Available from: https://www.dropbox.com/s/ao7c090p8bg1hk3/Lipton%20and%20Steinhardt%20-%20Troubling%20Trends%20in%20Machine%20Learning%20Scholarship.pdf?dl=0

25. Elish MC. Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction (We Robot 2016). 2016 [cited 2018 May 9];(We Robot 2016 Working Paper). Available from: https://papers.ssrn.com/abstract=2757236

26. Cath C, Wachter S, Mittelstadt B, Taddeo M, Floridi L. Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. Science and Engineering Ethics. 2017;1–24.

27. Marda V. India's Other Artificial Intelligence Challenge. Cath C, Wachter S, Mittelstadt B, Floridi L, editors. Philosophical Transactions of the Royal Society A. Forthcoming;(Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges).

28. Green B, Hu L. The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning. ICML 2018 debate papers [Internet]. 2018; Available from: https://www.dropbox.com/s/4tf5qz3mgft9ro7/Hu%20Green%20-%20Myth%20in%20the%20Methodology.pdf?dl=0

29. Pasquale F. The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge: Harvard University Press; 2016. 320 p.

30. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. 1 edition. New York: Crown; 2016. 272 p.

31. Noble SU. Algorithms of Oppression: How Search Engines Reinforce Racism. 1 edition. New York: NYU Press; 2018. 256 p.

32. Beer D. The social power of algorithms. Information, Communication & Society. 2017;20(1):1–13.