

MANAGING ARTIFICIAL INTELLIGENCE: POTENTIAL AND DIFFICULTIES FROM AN ETHICAL, LEGAL, AND TECHNICAL STANDPOINT

Priyanka Vijay Patil¹, Asavari Ajit Patil², Rupali Babaso Dhansare³

Assistant Professor^{1,2,3}

Department of Computer Application^{1,2,3}

Sanjay Ghodawat University^{1,2}, V.P. Institute Of Management Studies And Research Centre,
Sangli³

Abstract: The special issue titled "Governing Artificial Intelligence: Ethical, Legal, and Technical Opportunities and Challenges" has this piece as its introduction. Artificial intelligence (AI) is becoming increasingly prevalent in all facets of our lives, from the vital—such as banking, law enforcement, urban infrastructure, healthcare, and humanitarian aid—to the banal, such as dating. Artificial Intelligence (AI) has the potential to enhance economic, social, and human rights activities. This includes embodied AI in robotics and machine learning techniques. As AI becomes more prevalent in high-risk domains, there is an increasing need to create and regulate AI in a way that is transparent, equitable, and accountable. How and by what frameworks may this be accomplished? This is among the main issues covered in these eight writers' examination in-depth of the ethical, legal-regulatory, and technical issues raised by creating governance frameworks for AI systems in this special issue. In addition, it provides a brief summary of recent advancements in AI governance, outlines the extent to which the agenda for defining AI regulations, ethical frameworks, and technical techniques has been set, and offers some specific recommendations to forward the discussion of AI governance.

The topic issue "Governing Artificial Intelligence: Ethical, Legal, and Technological Opportunities and Challenges" includes this paper.

Keywords: artificial intelligence, law, ethics, technology, governance, culture

1. Introduction

Artificial intelligence (AI) is becoming more and more prevalent in all facets of our lives, from the vital—such as healthcare and humanitarian relief—to the banal—such as dating. Artificial Intelligence (AI) has the potential to improve economic, social welfare, and human rights exercise. This includes embodied AI in robotics and machine learning techniques. This new technology can be advantageous to the many areas mentioned. AI, however, has the potential to be abused or to act in unpredictable and dangerous ways. Thus, it is more important than ever to consider how the law, ethics, and technology should control AI systems. Alternatively, as Floridi [1] contends: "since the digital revolution affects our beliefs about values and objectives, socially desirable innovation that is both sustainable and acceptable in behavior, and governs all this has now become the fundamental issue' (p. 2).

AI systems are utilized in a wide range of applications; the majority of them use statistical learning techniques to identify patterns in massive data sets and generate predictions based on those patterns. As AI becomes more prevalent in high-risk fields, there is increasing demand to design and control AI in a transparent, equitable, and accountable manner. How can this be accomplished, and which frameworks can help? One of the main issues this special issue's writers address is this one. They provide in-depth studies of the ethical, legal-regulatory, and technical difficulties presented by creating governance regimes for AI systems.

AI systems are being used more and more by societies to handle risky, complex tasks like financial transaction management, patient diagnosis, and parole approval. This creates additional issues, such as liability for automated cars and the boundaries of the existing legal frameworks in dealing with 'big data's disproportionate impact' [2] or mitigating algorithmic damages [3], social justice challenges connected to automating law enforcement or social welfare [4], or online media consumption [5]. These important topics can only be adequately addressed from a multidisciplinary approach due to the broad impact of AI.

Eight original pieces from internationally renowned specialists in AI, computer science, data science, engineering, ethics, law, politics, robotics, and social sciences are included in this theme issue. The articles are updated versions of the papers that were given at three scheduled workshops. 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. "Ethical auditing for accountable automated decision-making," "Ethics & AI: responsibility & governance," and "Explainable and accountable algorithms" were the titles of the workshops. New suggestions for creating and promoting the ethical, legal, and technical governance of AI will be made in this special issue. It is concentrated on the examination of three distinct study domains:

- (1) Ethical governance: concentrating on the most relevant ethical questions that AI raises, addressing topics like justice, openness, and privacy (as well as how to react when the application of AI can result in widespread prejudice), the distribution of services and goods (the usage of AI by businesses, governments, and industries), as well as economic displacement (the moral response to the loss of jobs brought about by AI-based automation).
- (2) The notions of interpretability and explain ability are thought to be potential means of enhancing algorithmic accountability, transparency, and fairness. In Europe, for instance, there is discussion about the concept of a "right to explanation" for algorithmic choices. This privilege would give people the right to an explanation if an algorithm makes a decision regarding them (e.g. refusal of a loan application). This right isn't yet assured, though. Furthermore, it's unclear how the "perfect algorithmic explanation" would be defined and how AI systems might include these explanations.
- (3) Ethical auditing: accountability methods cannot rely exclusively on interpretability in the case of opaque and extremely sophisticated algorithmic systems. As potential remedies, auditing techniques that check for bias and damages in the inputs and outputs of algorithms are suggested as opposed to dissecting how the system works.

A rising corpus of literature addresses issues related to robotics and AI ethics [1,6–10], legislation [3,11–14] governing their effects [15], technological methods such as algorithmic impact evaluations [16–18], and establishing credibility through system validation [19]. These Law, ethics, and technology are the three guiding principles of AI governance that can work in concert [1]. Yet as Nemitz and Pagallo deftly point out in this issue, there is still disagreement over when one approach—or set of approaches—is most pertinent [13,17]. Globally, experts, governments, businesses, and civil society are arguing whether or not legal regulatory frameworks are necessary and whether or not ethical or technical approaches are sufficient. Even after those queries are resolved, the matter of how much our.

The impact of these technologies is sufficiently covered by the current ethical and regulatory frameworks. For example, Pagallo analyzes the discussion around the legal standing of embodied AI (robots) in the EU [13], highlighting this paradox. Veale and associates [3] assert that while European data protection offers strong guidelines, "many socio-technical difficulties created by machine learning and algorithmic systems more broadly are not entirely dealt with using the provisions of clauses in laws like the General Data Protection Regulation, which are the outcome of gradual changes in definitions and issues (p. 17). In particular, Winfield & Jirotko [9] address the function of technological standards in the flexible and moral regulation of robots and artificial intelligence.

Additionally, academia is arguing over its own strategy for AI governance. In a recent paper on 'troubling trends in machine learning scholarship', Lipton & Steinhardt [20], for example, warned against technical solutionism through the misuse of notions 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). In recent discussions, a number of scholars skillfully questioned the imaginaries that underpin data-driven technologies like artificial intelligence (AI) [21] and emphasized the dangers associated with using AI systems [22–24]. Applying these crucial lenses to the ethical, legal, and technical domains requires more effort—suggestions made for AI governance.

The sophisticated and complex state of the argument is reflected in the pieces in this special edition. The authors also demonstrate how some of the suggested legal governance alternatives have too narrow a focus. In addition to showing that specific moral answers have conceptual impreciseness and absence of procedures for enforcement. Similarly, some technology approaches have the potential to reduce transparency to a box-ticking exercise or oversimplify complex social issues like justice. Therefore, the essays in this special issue not only offer suggestions for additional ethical, legal, and technical improvements, but they also provide a critical evaluation of the current state of AI governance. By doing this, the authors emphasize how crucial it is to take into account who is in charge of AI governance and what these people and organizations stand to individuals and organizations stand to gain. Because, as stated by Harambam et al. [5, p. 1],

"technology is, after all, always the product of our making, including the course it may take," rather than an unstoppable or uncontrollable force of nature. Even with artificial intelligence.

There are undoubtedly unanswered concerns about what constitutes appropriate AI governance [2, 25–27]. Political organizations from all around the world, such as the UK [28], South Korea [29], India [30], and the Mexican government [31], are currently debating these issues. well as the European Commission [32]. Our goal is to influence these discussions in some way with the pieces in this special edition. A brief synopsis of current advancements in AI governance, including agenda-setting processes for establishing technical approaches, ethical frameworks, and regulations pertaining to AI, will be provided in order to contextualize the various articles.

2. Establishing the AI governance agenda

Academics and policymakers are finding it difficult to stay up to date with the volume of papers, guidelines, laws, and technical standards being created on AI governance. At least a dozen nations unveiled new AI policies in the first half of 2018 alone [33], with several promising government support of up to 1.8 billion [34]. In the meanwhile, the industry is creating its own AI principles or initiating multistake holder projects to create best practices. Through direct involvement or lobbying efforts, they are also active in the development of regulations pertaining to AI.

While these corporate initiatives are commendable, it is crucial to consider them in the context of three key questions. Who establishes the agenda for AI governance first? Secondly, what cultural logic does that agenda actualize, and thirdly, who stands to gain from it? The answers to these questions are crucial because they expose the dangers of allowing businesses to set the agenda and expose myopic places in ongoing studies.

There is excellent work on the problematic developments in machine learning research that arise from the confusion of simple statistics with complex social notions [20, 35]. Likewise, numerous writers emphasize how the unregulated application of "black box" systems in banking [36], education, Search engines [38], social welfare [4], and criminal justice [37] can all have negative consequences. The goal of Beer's [39] argument is to center it around the "social power of algorithms." According to him, "part of the discursive reinforcement of particular norms, approaches, and modes of reasoning" is how the cultural concept of algorithm functions (p. 11). As previously noted, the ways in which AI systems are perceived and understood [21] as well as their operational mechanisms influence AI governance. The following paragraphs will draw attention to a few issues and encourage a closer examination of the cultural logic put out by actively influencing the conversation.

The USA is home to a large number of the top AI companies in the world. The degree to which AI systems replicate societies in the image of US culture and the inclinations of US tech giants is an evident cause for concern. Programming AI doesn't always require a lot of resources. Its stored data accounts for a large portion of its worth. As a result, a small number of American businesses drive the majority of technological innovation.² Considering that these businesses are leading the way in a number of regulatory measures, it is crucial to make sure that this specific worry is not

made worse. The rest of the world is not going to automatically align with an agenda dictated by corporate needs in America.

AI systems are frequently described as "black boxes" [36], which are extremely complicated and challenging to understand [23]. Kroll [19] shows that these arguments conceal that algorithms are fundamentally understandable. He contends that "instead of disregarding systems that result in negative results as Since the use of subpar technology is essentially mysterious and so unmanageable, we should just call it what it is: malpractice carried out by a system's controller. (p. 5). However, the intimate participation of the AI sector in regulation and policy-making is frequently justified by the cultural logic of the "complex inscrutable" technology [40]. The industry participants in these policy procedures typically comprise the same elite group that is in charge of online marketing and data collection. This isn't an isolated incident. Businesses as Large amounts of data can be gathered by Google, Facebook, and Amazon, which can be utilized to advance new AI-based services. Thus, the "turn to AI" legitimizes the presence of large corporations in regulatory procedures while also strengthening their position in the market.

The influence that corporations have over AI policy is a related worry. They function as semi-co-regulators at times. For instance, the CEO of Facebook testified regarding his company's involvement in the data leak during a joint hearing of the US Senate Commerce and Judiciary Committees following the Cambridge Analytica incident. Several Senators specifically requested him [41] to give examples of the kinds of regulations his company should be subject to during the session. Similarly, a High-Level Expert Group on AI was recently established by the European Commission [42]. The group's task is to collaborate with the Commission on putting a European AI strategy into action.

The 52 members of the organization come from a variety of backgrounds, and while not all links are obvious, it seems that about half come from the industry, followed by academics (17 members) and civil society (just 4 members). In this issue, Marda emphasizes how crucial it is to make sure Civil society has an equal voice in the development of AI governance regimes since it is frequently the group most impacted by AI systems. She demonstrates how, at the expense of social and ethical issues, the contemporary discourse in India is primarily cantered on industrial and governmental concerns as well as the objectives of innovation and economic growth [27].

Nemitz [17] also emphasizes how a small number of companies have a significant amount of influence in the AI industry. In this issue, he writes: "The critical investigation into how emerging technologies like artificial intelligence relate to human rights, democracy, and the rule of law must thus start from a comprehensive examination of contemporary business models and technology, including the concentration of political, economic, and technological power in the hands of the "frightful five," who are central to the development and systems integration of artificial intelligence into services that can be profitably offered. The development of numerous significant international efforts on AI and ethics is another way that industry impact may be seen.

Establishing open norm-setting forums with the goal of addressing AI governance through the development of technical standards, moral guidelines, and professional conduct norms has several benefits. However the suggested fixes should do more to go above and beyond existing voluntary

ethical frameworks or strictly technical definitions of openness, accountability, and justice. The several pieces in this edition make it abundantly evident why the issues of strict regulation and the internet's attention-advertising-based business model must be addressed going forward. A comprehensive approach to addressing these concerns is necessary if we are to take AI governance seriously.

3. Conclusion

The argument made in this article should not be seen as discounting the efforts made by business or the applicability of the current technological, ethical, and regulatory frameworks for AI governance. Instead, there is a lot to gain from this ongoing work—but only if we carefully consider its objectives, implications, and methodology. The fundamental goals of AI governance solutions must be critically questioned, as must any unintended cultural side effects. This is particularly true when it comes to establishing the legitimacy of norms pertaining to ethics, standards, and regulations that are developed by the private sector. Similarly, we need to continue to be aware of the issues that are either entirely unaddressed or only partially addressed by terms like accountability, openness, and justice. What is not talked about when concentrating on these issues? Are we supposing that these well-known acronyms inherently cover concerns related to AI and equality, social justice, or human rights? Or do the groups promoting the agenda consider these issues outside of their purview? It is important to ask these difficult questions as these ideas are finding their way into global legislative initiatives [43].

The writers of this special issue tackle these many challenging topics with skill. It is evident from the essays that the writers are dissatisfied with AI governance as it is at the moment. Nemitz, for example, is an advocate of developing a new corporate and technological culture.

Human rights, the rule of law, and democratic ideals served as the foundation for development [17]. Pagallo emphasizes the value of pragmatism and the use of legal experimentation techniques to explore novel theories of accountability and liability [13]. According to Veale et al., "enabling users to deploy local personalization tools might balance power relations in relation to large firms hoarding personal data" [3, p. 5]. They investigate how machine learning models might be regarded as personal data under European data protection law. Strong ethical standards are just the first step, according to Winfield and Jirotko, and additional work needs to be done to ensure accountability and execution. When the rubber meets the road—or rather, the robot—that is when the true test of proper governance of AI systems occurs.

To reduce the risks associated with attempts to achieve algorithmic transparency, as well as to give people more control over the algorithms used in the news industry, Harambam et al. investigate the concept of "voice" [5]. Here and in earlier articles [26], the editors contended that When regulating AI, it's critical to make sure that all stakeholders are fairly represented. More non-US-led projects are also required, such as the Council on Europe's Expert Committee on AI and Human Rights and the Europe-based AI4People. To reduce the risks associated with attempts to achieve algorithmic transparency, as well as to give people more control over the algorithms used in the news industry, Harambam et al. investigate the concept of "voice" [5]. Here and in earlier articles [26], the editors contended that When regulating AI, it's critical to make sure that all stakeholders

are fairly represented. More non-US-led projects are also required, such as the Council on Europe's Expert Committee on AI and Human Rights and the Europe-based AI4People. More European-led initiatives are vital, but we also need to consider the concerns of the Global South. The reason these voices are particularly important is highlighted in Marda's article about India [27]. Likewise, it's critical to move behind the language of responsibility, justice, and openness to Determine what other core values ought to be mentioned. Nemitz, Florida, and Marda, for instance, make the case that human rights principles should be included [1, 17, 27].

Ultimately, the critical viewpoints presented in this special issue shed light on the subtleties of the discussion around artificial intelligence, ethics, technology, and the law and open the door to a more expansive and inclusive agenda for AI governance. Alternatively, as Kroll reminds us, "power dynamics between actors that exist independent of the technical tools in use generally result in opacity in socio-technical systems." Software systems are no different from other artifacts in that they cannot be fully understood in isolation from their human context [19, p. 11]. The writers' careful consideration of the subjects covered in this special issue is greatly appreciated by the editors. Their work serves as an excellent example of the kind of multidisciplinary study that is required. A few different issues were attempted to be highlighted in this introduction piece. covered by the writers, however, the condensed summaries that are provided fall short of fully capturing the nuanced and compelling arguments put forth in each of the individual articles. The three main subjects of this special issue—ethical auditing, explainability and interpretability, and ethical governance—are all represented in the articles, which also offer a critical evaluation of AI governance as it stands today. Throughout this special issue the reader is invited to, as Floridi argues, resist the distracting narrative that 'digital innovation leads, and everything else lags behind, or follows at best: business models, working conditions, standards of living, legislation, social norms, habits, expectations and even hope' [1, p. 2].

References

1. Floridi L. 2018 Soft ethics, the governance of the digital and the General Data Protection Regulation. *Phil. Trans. R. Soc. A* 376, 20180081. (doi:10.1098/rsta.2018.0081)
2. Barocas S, Selbst AD. 2016 Big data's disparate impact. *Cal. L. Rev.* 104, 671.
3. Veale M, Binns R, Edwards L. 2018 Algorithms that remember: model inversion attacks and data protection law. *Phil. Trans. R. Soc. A* 376, 20180083. (doi:10.1098/rsta.2018.0083)
4. Eubanks V. 2018 Automating inequality: how high-tech tools profile, police, and punish the poor. New York, NY: St. Martin's Press.
5. Harambam J, Helberger N, van Hoboken J. 2018 Democratizing algorithmic news recommenders: how to materialize voice in a technologically saturated media ecosystem. *Phil. Trans. R. Soc. A* 376, 20180088. (doi:10.1098/rsta.2018.0088)

6. Floridi L, Taddeo M. 2016 What is data ethics? *Phil. Trans. R. Soc. A* 374, 20160360. (doi:10.1098/rsta.2016.0360)
7. Ananny M. 2016 Toward an ethics of algorithms: convening, observation, probability, and timeliness. *Sci. Technol. Hum. Values* 41, 93–117. (doi:10.1177/0162243915606523)
8. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. 2016 The ethics of algorithms: mapping the debate. *Big Data Soc.* 3, 2053951716679679. (doi:10.1177/2053951716679679)
9. Winfield AFT, Jirotko M. 2018 Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Phil. Trans. R. Soc. A* 376, 20180085. (doi:10.1098/rsta.2018.0085)
10. Taddeo M, Floridi L. 2018 How can AI be a force for good: an ethical framework will help harness the potential of AI while keeping humans in control. *Science* 361, 751–752. (doi:10.1126/science.aat5991)
11. Wachter S, Mittelstadt B, Floridi L. 2017 Why a right to an explanation of automated decisionmaking does not exist in the general data protection regulation. *Int. Data Privacy Law* 7, 76–99. (doi:10.1093/idpl/ix005)
12. Edwards L, Veale M. 2018 Enslaving the algorithm: from a ‘right to an explanation’ to a ‘right to better decisions’? *IEEE Security Privacy Mag.* 16, 46–54. (doi:10.1109/MSP.2018.2701152)
13. Pagallo U. 2018 Apples, oranges, robots: four misunderstandings in today’s debate on the legal status of AI systems. *Phil. Trans. R. Soc. A* 376, 20180168. (doi:10.1098/rsta.2018.0168)
14. Mendoza I, Bygrave LA. 2017 The right not to be subject to automated decisions based on profiling. University of Oslo Faculty of Law Research Paper, Norway.
15. Wachter S, Mittelstadt B, Floridi L. 2017 Transparent, explainable, and accountable AI for robotics. *Sci. Robot.* 2, eaan6080. (doi:10.1126/scirobotics.aan6080)
16. Selbst AD. 2017 Disparate impact in big data policing. *Georgia Law Rev.* 52, 109.
17. Nemitz P. 2018 Constitutional democracy and technology in the age of artificial intelligence. *Phil. Trans. R. Soc. A* 376, 20180089. (doi:10.1098/rsta.2018.0089)
18. AI NOW. 2018 Algorithmic impact assessments: a practical framework for public agency accountability. New York: NY: AI NOW.

19. Kroll JA. 2018 The fallacy of inscrutability. *Phil. Trans. R. Soc. A* 376, 20180084. (doi:10.1098/rsta.2018.0084)
20. Lipton ZCL, Steinhardt J. 2018 Troubling trends in machine learning scholarship. ICML 2018 debate papers. See <https://www.dropbox.com/s/ao7c090p8bg1hk3/Lipton%20and%20Steinhardt%20-%20Troubling%20Trends%20in%20Machine%20Learning%20Scholarship.pdf?dl=0>.
21. Elish MC, Boyd D. 2017 Situating methods in the magic of big data and AI. *Commun. Monogr.*85, 57–80.
22. Citron DK. 2007 Technological due process. Report no.: ID 1012360. Rochester, NY: Social Science Research Network.
23. Burrell J. 2016 How the machine ‘thinks’: understanding opacity in machine learning algorithms. *Big Data Soc.* 3, 2053951715622512. (doi:10.1177/2053951715622512)
24. Crawford K, Calo R. 2016 There is a blind spot in AI research. *Nature News* 538, 311. (doi:10.1038/538311a)
25. Elish MC. 2016 Moral crumple zones: cautionary tales in human-robot interaction (We Robot 2016). [cited 2018 May 9]; (We Robot 2016 Working Paper). See <https://papers.ssrn.com/abstract=2757236>.
26. Cath C, Wachter S, Mittelstadt B, Taddeo M, Floridi L. 2017 Artificial intelligence and the ‘good society’: the US, EU, and UK approach. *Sci. Eng. Ethics* 24, 505–528.
27. Marda V. 2018 Artificial intelligence policy in India: a framework for engaging the limits of data-driven decision-making. *Phil. Trans. R. Soc. A* 376, 20180087. (doi:10.1098/rsta.2018.0087)
28. House of Commons. 2017 Algorithms in decision-making. House of Commons Science and Technology Committee. See <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf>.
29. Korea Times. 2018 Korea to nurture artificial intelligence, robot as new cash cow. See https://www.koreatimes.co.kr/www/news/biz/2016/10/123_217161.html.
30. Times of India. 2018 NITI Aayog releases strategy on artificial intelligence, identifies 5 focus areas. See http://timesofindia.indiatimes.com/articleshow/64452469.cms?utm_

source=contentofinterest&utm_medium=text&utm_campaign=cppst.

31. Mexican Government. 2018 Estrategia de Inteligencia Artificial MX 2018. See <https://datos.gob.mx/blog/estrategia-de-inteligencia-artificial-mx-018?category=noticias&tag=nula>.

32. European Commission. 2018 Artificial intelligence: Commission outlines a European approach to boost investment and set ethical guidelines. See http://europa.eu/rapid/pressrelease_IP-18-3362_en.htm.

33. Dutton T. 2018 Politics of AI, an overview of national AI strategies. See <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>.