AI vs. Human: Reconciling Autonomy and Governance in Decision-Making

Dowson Liu[1], Xue Li[2*]

[1]RapidsDB, Inc., San Jose, USA
[2]eXomAI Limited, Hong Kong, China
**Corresponding Author. Email: xueli@scholar42.com*

Abstract. This report addresses the escalating conflict between decision-making processes augmented by artificial intelligence and the safeguarding of human autonomy. The focal point of our investigation rests on the ethical ramifications arising from the increasing integration of AI across crucial sectors. The analysis considers applications in areas such as healthcare, finance, and law, where AI enhances efficiency and scalability but also heightens risks of bias, opacity, and reduced accountability. Further, we delve into the concept of AI governance as a mechanism to ensure sustained societal oversight and control. Governance is approached not only as regulation but as an adaptive framework that incorporates transparency, inclusivity, and continuous monitoring to address emerging challenges. The objective is to elucidate the inherent challenges in harmonizing the efficiency and scalability offered by AI systems with the imperative of preserving human agency and ethical considerations in decision-making frameworks. By synthesizing sectoral impacts, ethical concerns, and governance strategies, the report highlights both the transformative promise of AI and the urgent need for policies that safeguard fairness and human values. The report provides an overview of the current landscape, highlighting potential risks and proposing avenues for future research and policy development in the realm of AI governance.

Keywords: AI Governance, Human Autonomy, Ethical Implications, Decision-Making, AI Integration

1. Introduction

Artificial intelligence (AI) is rapidly transforming decision-making processes across a multitude of sectors, from healthcare and finance to transportation and criminal justice. This integration promises increased efficiency, accuracy, and scalability. However, it also introduces significant challenges concerning human autonomy and control. The increasing reliance on AI systems to inform or even automate critical decisions raises fundamental questions about accountability, transparency, and the potential for bias or unintended consequences.

This report provides an overview of the current landscape of AI-assisted decision-making, focusing on the tension between leveraging the benefits of AI and safeguarding human autonomy. It outlines key challenges, such as algorithmic bias and the ’black box’ nature of some AI models, that impede effective AI governance. Furthermore, it explores opportunities for developing robust

governance frameworks that promote responsible AI development and deployment, ensuring that AI systems are aligned with human values and societal goals. The objective is to stimulate a constructive dialogue on how to best navigate the ethical and practical considerations surrounding AI in decision-making, paving the way for a future where AI serves as a beneficial tool that complements, rather than compromises, human judgment and autonomy.

2. AI's impact on decision-making

Artificial intelligence is rapidly transforming various sectors, bringing both unprecedented opportunities and potential challenges to decision-making processes. The increasing reliance on AI in key industries such as healthcare, finance, and law has the potential to revolutionize these fields by enhancing efficiency, accuracy, and scalability. However, this transformative technology also introduces a range of ethical concerns and practical risks, particularly regarding fairness, transparency, and accountability. This section delves deeper into AI's current applications in healthcare, finance, and law, highlighting both the substantial benefits and the critical risks associated with these technologies.

2.1. Healthcare

AI is increasingly used in medical diagnosis, offering the potential for faster and more accurate identification of diseases. Foundation models like RETFound, trained on vast datasets of retinal images, demonstrate the ability to diagnose sight-threatening eye diseases and predict systemic disorders with fewer labeled data [1]. The prospect of generalist medical AI (GMAI) promises even more versatile applications, capable of interpreting diverse medical modalities and providing comprehensive outputs [2]. However, concerns remain regarding algorithmic bias, particularly the underdiagnosis of underserved patient populations, which could exacerbate existing healthcare disparities [3]. Algorithmic bias can arise from systemic inequalities in dataset curation and unequal access to research participation [4]. AI is also making inroads into medical robotics, with potential for autonomous robots to perform diagnostic imaging, remote surgery, and rehabilitation tasks [5], and to improve surgical care across pre-operative, intra-operative, and post-operative settings [6].

2.2. Finance

In finance, AI algorithms are employed for tasks such as fraud detection, risk assessment, and algorithmic trading. While these applications can enhance efficiency and profitability, they also pose risks related to market manipulation and algorithmic bias. Algorithmic bias, as noted in [7], can stem from the choice of seemingly effective proxies that mask underlying inequalities, leading to discriminatory outcomes in financial services. The application of Judea Pearl's work on causality [8] could provide a deeper understanding of the causal connections between financial variables and help to devise less biased models.

2.3. Law

AI is being integrated into the legal field for tasks such as legal research, contract analysis, and predictive policing. Although AI-powered tools can improve efficiency and accuracy, they also raise concerns about transparency, accountability, and fairness. The use of machine learning algorithms in criminal justice, for instance, may perpetuate existing societal biases, leading to inequitable sentencing and policing practices. As highlighted by [9], semantics derived from language corpora

can contain human-like biases that are then amplified by machine learning models. Addressing such biases requires careful consideration of fairness definitions and the implementation of fairness-enhancing mechanisms [10-12].

In summary, AI's impact on decision-making is multifaceted, offering significant potential benefits but also presenting risks related to bias, inequity, and lack of transparency. While AI systems can improve efficiency and accuracy in healthcare, finance, and law, their deployment must be carefully managed to avoid exacerbating existing inequalities. Ensuring fairness, transparency, and accountability in AI systems is crucial for their responsible use. A focus on data diversity, causal understanding, and fairness-aware design will be necessary to harness AI's power ethically and responsibly, ensuring that it complements human decision-making while mitigating risks of bias and inequity.

3. Ethical considerations

AI's increasing role in decision-making presents a range of ethical considerations, particularly when these decisions have the potential to alter lives significantly [13]. These ethical concerns span multiple dimensions, encompassing fairness, accountability, and transparency [14]. The potential for algorithmic bias to disproportionately affect vulnerable populations represents a critical challenge [3, 15]. As AI becomes more embedded in systems that impact healthcare, law enforcement, hiring, and other areas, these biases can exacerbate existing societal inequalities, further marginalizing already disadvantaged groups.

Algorithmic bias, as evidenced in various AI applications, underscores the necessity for careful scrutiny and mitigation strategies [10, 16]. The sources of these biases are diverse, ranging from skewed training data to inherent limitations in the algorithms themselves [9, 17]. These biases can result in AI systems that perpetuate and even amplify existing societal inequalities, particularly concerning race, gender, and socioeconomic status [18, 19]. For instance, AI systems used in predictive policing or hiring processes may inadvertently reinforce stereotypes, undermining efforts toward equality and fairness.

The increasing reliance on AI in high-stakes decision-making environments calls for frameworks that ensure accountability for these systems. To address these concerns, several guidelines and frameworks have been proposed to enhance transparency and accountability in AI interventions. For example, the CONSORT-AI extension offers reporting guidelines for clinical trials involving AI, promoting clearer descriptions of AI interventions and analyses of error cases [20]. Such frameworks can help ensure that AI systems are not treated as "black boxes," but are instead subject to meaningful scrutiny. Similarly, there's a growing emphasis on transparency in data sharing and open science collaboration to promote more robust and equitable AI systems [21, 22]. This push for transparency is critical in fostering trust in AI systems and ensuring that they are designed and deployed with fairness in mind.

In considering biased data, Ferryman et al. suggest thinking of clinical data as artifacts that reflect the values and patterns of inequity that exist in society [17]. This perspective encourages a more thoughtful examination of the data used to train AI systems, prompting questions about whose voices are represented and whose are excluded. This approach also highlights the importance of involving diverse stakeholders in the creation and evaluation of AI systems to ensure that they reflect a broad range of perspectives and needs.

Despite the growing awareness of these ethical challenges, it remains imperative to develop comprehensive ethical guidelines for AI, focusing on responsible use, accurate and reliable information exchange, protection of patient privacy, and empowerment of individuals in their

decision-making processes [14]. It is not enough to simply acknowledge the ethical challenges posed by AI; there must be concrete strategies to address them and prevent harm. Further research is needed to determine the extent to which such parameters affect the fairness and reliability of AI systems [23]. By addressing these ethical considerations proactively, AI can be harnessed as a tool for societal benefit while safeguarding against potential harms. Additionally, creating a culture of ethical responsibility within AI development teams is essential, ensuring that fairness and accountability are prioritized throughout the lifecycle of these technologies. In this way, AI can fulfill its promise as a force for good without exacerbating existing social inequalities.

4. AI governance and societal control

The transformative power of Large Generative AI Models (LGAIMs) necessitates a critical examination of AI governance and the means by which society can exert control over their decision-making processes [24]. The EU and other regions are increasingly focused on AI regulation, but these efforts must adapt to the specific attributes of LGAIMs [24]. One approach involves defining obligations across the AI value chain, encompassing developers, deployers, and users [24]. This ensures that accountability is maintained throughout the AI lifecycle, from development to deployment, and that all actors are held responsible for the ethical implications of their contributions.

Trustworthy AI hinges on lawfulness, ethical considerations, and robustness [25]. Díaz-Rodríguez et al. [25] propose a holistic vision incorporating ethical principles, philosophical considerations, risk-based regulation, and key requirements such as human agency, privacy, transparency, and accountability. Their framework calls for a balance between technological progress and safeguarding fundamental human rights. The increasing prevalence of AI in critical sectors requires that ethical regulations include foresight methodologies to identify potential risks and unwanted consequences [26]. These methodologies should be embedded in regulatory processes to ensure that AI systems are not only effective but also aligned with societal values and the well-being of the public.

In healthcare, for example, ensuring the safe and timely translation of AI research into validated systems is crucial [27]. This involves robust clinical evaluation, intuitive metrics, and addressing algorithmic bias, brittleness, and interpretability [27]. Such evaluations are necessary to ensure that AI systems are reliable and that they do not introduce or exacerbate disparities in healthcare delivery, particularly for vulnerable populations.

The roles of government, industry, and civil society are paramount in shaping AI's future. Hacker, Engel, and Mauer [24] argue for a three-layered obligation system: minimum standards, high-risk obligations for specific use cases, and collaborative efforts across the AI value chain. This multi-layered approach ensures that regulation is tailored to the risks associated with different AI applications, with a particular focus on high-risk sectors. Regulations should focus on concrete high-risk applications, emphasizing transparency and risk management [24]. These regulations should also prioritize accountability mechanisms that ensure AI's benefits are equitably distributed, especially in sectors with significant societal impact.

Considering Indigenous data sovereignty, as discussed by Kukutai [28], also adds a critical dimension to the responsible governance of AI. Indigenous communities, whose data are often used without consent or adequate benefits, must be empowered to maintain control over their data and ensure its ethical use in AI systems. The inclusion of such considerations highlights the importance of respecting diverse cultural values and legal frameworks in AI governance.

5. Conclusion

In summary, this report has explored the crucial intersection of artificial intelligence and decision-making, underscoring the necessity of carefully navigating the integration of AI technologies within various sectors. The core findings highlight the dual potential of AI to enhance efficiency and accuracy while simultaneously posing significant ethical and societal challenges.

The exploration has reinforced the importance of proactively addressing concerns related to bias, transparency, and accountability in AI systems. As AI's role in decision-making expands, it is imperative to establish clear guidelines and frameworks that prioritize fairness, equity, and respect for human autonomy. These frameworks should not only govern the development and deployment of AI but also ensure continuous monitoring and evaluation to mitigate unintended consequences.

Ultimately, the successful adoption of AI hinges on striking a delicate equilibrium between leveraging its capabilities and safeguarding fundamental human values. This requires a concerted effort from policymakers, researchers, and industry stakeholders to foster a responsible and human-centered approach to AI innovation. By prioritizing ethical considerations and promoting transparent governance, we can harness the transformative power of AI to benefit society as a whole.

References

[1] Zhou Y, Chia MA, Wagner S, Ayhan MS, Williamson DJ, Struyven R, et al. A foundation model for generalizable disease detection from retinal images. Nature. 2023

[2] Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023

[3] Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nature Medicine. 2021

[4] Arora A, Alderman J, Palmer J, Ganapathi S, Laws E, McCradden MD, et al. The value of standards for health datasets in artificial intelligence-based applications. Nature Medicine. 2023

[5] Yip MC, Salcudean SE, Goldberg K, Althoefer K, Menciassi A, Opfermann JD, et al. Artificial intelligence meets medical robotics. Science. 2023

[6] Varghese C, Harrison EM, O'Grady G, Topol EJ. Artificial intelligence in surgery. Nature Medicine. 2024

[7] Obermeyer Z, Powers BW, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019

[8] Pearl J. Causality. 2009

[9] Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. Science. 2017

[10] Mehrabi N, Morstatter F, Saxena NA, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Computing Surveys. 2021

[11] Pessach D, Shmueli E. A review on fairness in machine learning. ACM Computing Surveys. 2022;

[12] Benjamin R. Assessing risk, automating racism. Science. 2019

[13] Mittelstadt B, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. Big Data & Society. 2016

[14] Wang C, Liu S, Yang H, Jiu-lin G, Wu Y, Liu J. Ethical considerations of using ChatGPT in health care. Journal of Medical Internet Research. 2023

[15] Birhane A. Algorithmic injustice: A relational ethics approach. Patterns. 2021

[16] Ferrara E. Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. Sci. 2023

[17] Ferryman K, Mackintosh M, Ghassemi M. Considering biased data as informative artifacts in AI-assisted health care. New England Journal of Medicine. 2023

[18] Cerdeña JP, Plaisime MV, Tsai J. From race-based to race-conscious medicine: How anti-racist uprisings call us to act. The Lancet. 2020

[19] Didier C, Duan W, Dupuy J, Guston DH, Yong-mou L, Cerezo JAL, et al. Acknowledging AI's dark side. Science. 2015

[20] Liu X, Rivera SC, Moher D, Calvert M, Denniston AK, Chan A, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. Nature Medicine. 2020
[21] Nosek BA, Alter G, Banks GC, Borsboom D, Bowman S, Breckler SJ, et al. Promoting an open research culture. Science. 2015
[22] Fiume M, Cupák M, Keenan S, Rambla J, Torre S de la, Dyke SOM, et al. Federated discovery and sharing of genomic data using beacons. Nature Biotechnology. 2019
[23] Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. Social Science & Medicine. 2021
[24] Hacker P, Engel A, Mauer M. Regulating ChatGPT and other large generative AI models. 2022 ACM Conference on Fairness, Accountability, and Transparency. 2023
[25] Díaz-Rodríguez N, Ser JD, Coeckelbergh M, Prado ML de, Herrera–Viedma E, Herrera F. Connecting the dots in trustworthy artificial intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion. 2023
[26] Fahrenkamp-Uppenbrink J. An ethical way forward for AI. Science. 2018
[27] Kelly C, Karthikesalingam A, Suleyman M, Corrado GS, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine. 2019
[28] Kukutai T. How indigenous communities in new zealand are protecting their data. Science. 2024