



The Ethics of Artificial Intelligence in Society

Kabiga Chelule Kwemoi

Faculty of Engineering Kampala International University Uganda

ABSTRACT

This paper examines the ethical implications of artificial intelligence (AI) as it becomes increasingly integrated into diverse sectors, including healthcare, finance, and employment. By addressing pressing ethical concerns—such as accountability, transparency, bias, privacy, and the social consequences of automation—the paper argues for the need to adapt existing ethical frameworks and, when necessary, create new ones. The analysis discusses the potential for AI to reinforce societal biases, the privacy challenges associated with data-intensive technologies, and the transformative impact of AI on employment. The paper emphasizes the importance of fostering a globally inclusive discourse to guide AI's development responsibly, prioritizing fairness, accountability, and public trust. Ultimately, the ethical integration of AI demands a collaborative approach, ensuring that AI serves as a tool to promote societal well-being rather than exacerbate inequalities.

Keywords: Artificial Intelligence, AI Ethics, Transparency, Bias in AI, Data Privacy, Accountability.

INTRODUCTION

Artificial Intelligence (AI) is increasingly finding its way into many sectors, such as health, finance, work, and automated driving. Not only are AI experts but also the general public, interested in potential ethical challenges. The boundaries of the ethics of AI are still blurred, but several pressing questions and challenges arise: Could AI influence human behavior and societal norms? How accountable are system designers and programmers for unwanted consequences? What are discriminatory biases, and how does AI influence such biases in society? Should AI be transparent? How can we achieve integration into society of an AI system that is explainable? These questions are not new, but they pose challenges in new ways – in particular, in how they enable us to think about accountability and transparency as fundamental components of the ethics of AI [1, 2]. One of the major aims of this paper is to show that AI is not the end of ethical considerations; AI is the beginning and an immediate case that forces us to revisit already existing ethics and, if necessary, replace them with new ones. This is why we need public engagement on these questions and global discourse – to better understand what kind of societies we want to be. Here we consider AI as a possible force for technological singularity. In the paper, AI refers to the narrow or weak AI, which allocates a lot of resources into programming systems to reason about and solve specific problem tasks. The review does not focus on strong AI, the hypothesis that a future super-intelligent AI could become more powerful than human brains and lead to the rise of superintelligence and even a technological singularity [3, 4].

Ethical Considerations in AI Development

The development of artificial intelligence (AI) technology is becoming an increasingly interdisciplinary collaboration of computer science, cognitive psychology, and philosophy, with significant contributions from other areas. Researchers and social scientists are adding their expertise to the discussion of the ethical issues surrounding this new field. Application areas range from control systems for industry, commerce, and the military to educational software for children and adults, and include legal, medical, and financial domains. Therefore, it is necessary to develop and implement technological tools and procedures that defy the unethical use of brain-computer interface systems. The following section outlines some of the key ethical considerations in the development of AI systems [5, 6]. There are many areas of ethical consideration in the development of these types of agents, which systems designers should

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

consider. These range from whether such systems should be designed at all, through consideration of what societal benefits they may achieve, and how the benefits might be maximized and reported. We hope that some currently unanswered questions about the design and development of these systems can be addressed. It is now difficult to develop a fixed set of principles to be generally followed in AI, if only because considerations of what used to be considered "fixed" values change and are not universally accepted [7, 8].

Bias and Fairness in AI Systems

One of the major issues associated with AI systems is bias. The design and execution of most AI systems come from humans. These humans may inadvertently encode biases into the algorithms and data that are used to design the AI systems. As a result, decisions and suggestions made by AI technologies may reinforce and even exacerbate existing biases, negatively affecting individuals based on their race, gender, sexual orientation, and other demographic characteristics. This may result in AI-based decisions being systematically unfair and perpetuating social inequalities. The consequences of unfair AI systems that share biases with current decision-making systems can be serious. There have been many cases of biased AI system software, including those that have affected hiring, lending, adjudication, judicial decisions, and others [9, 10]. To address issues of bias and to build fair AI systems, one must first measure and aggregate individual biases. Many mitigations for addressing bias are based on removing the relationship between sensitive attributes and either the predictions made by the AI system or the decisions that are ameliorated in response to these predictions. Addressing bias is a fundamentally ethical and political matter, and there will likely not be a single, universally agreed-upon metric for fairness. Rather, the choice of fairness metric should depend on the context in which the AI system is being used. When designing AI systems, people should consider different dimensions of fairness and work together to make equitable choices. The above suggests that the principles, as well as the mechanics, of the fairness of artificial intelligence, are important for the evaluation of its ethical impact. With AI systems already prevailing, questions of unbiasedness are effectively also about questions of moral permissibility. Accordingly, unveiling AI systems to the public can be especially beneficial: by doing so, organizations can increase transparency and foster public trust [11, 12].

Privacy and Data Protection

Artificial intelligence applications are often fueled or supported by vast amounts of data. This typically includes personal information about individuals, including their identity, location, preferences, behavior, social relations, etc. This practice raises profound questions about society's framing of privacy and the ethical boundaries of data usage. In a digital society, where even daily activities and movements are heavily coupled with digital transactions and documentation, the stakes for privacy have never been higher. Personal data, together with the implications of its collection, usage, and storage in profiles about each individual, is tied to several human rights and fundamental principles of the regulatory framework. These include protection from discrimination, economic and consumer harm, identity theft, malicious communications, protection against socio-political manipulation and abuse, or algorithmic violence. Data protection laws generally see the balance between prohibition, restrictions, and permission to be contingent on the data subject's rights and liberty at stake rather than on the innovation-led public good on the other side of the balance [13, 14]. The rights-based approach at the core of privacy laws has more recently been engineered in the European General Data Protection Regulation and its provision for Privacy by Design. This latest generation of data protection laws focuses on leading the data controllers in a new process-driven manner, to ensure the alignment of data processing to the individual rights more generally. Given the risks involved, GDPR also mandates privacy impact assessments for high-risk processing activities. The collection and use of personal or personal-like data are now subject to very detailed requirements imposed by a multi-tier regulation whereby the basic principles are complemented by physical security, data and metadata protection of the collection and transfer, storage, processing, and retention, reducing the potential for harm that can be caused to society by the use of new technologies such as AI and machine learning algorithms. In some cases, technology provides the required level of protection of personal data. For personal data processing operations, such as in health, where the risk can affect the rights and freedoms of individuals, GDPR also requires data to be stored in an encrypted manner, which means that on top of being difficult to hack, the data is also unreadable as an added security feature. At a process level, the new European data protection regulation requires that organizations obtain the consent of users when collecting and processing personal data. This requirement can also be seen as an ethical principle. This is primarily because being subject to someone else's decision can represent a challenge for autonomy. At a metadata level, the requirement for anonymization in AI and

machine learning is essential to reduce the risk of sharing individual personal information. Therefore, any data shared for AI processing should be anonymous such that the record can no longer be attributed to an individual, or to reduce the probability of re-identification of the individual, data minimization should be achieved by not retaining data that is not necessary for the achievement of the specific purpose. It should also not be so high as to effectively render the data in question useless for various purposes of AI applications [15, 16].

AI And Employment Impact

Artificial intelligence is coupled with a potential impact on employment and the workforce. The adoption of AI technologies might result in developments such as job displacement due to industrial automation processes. In contrast, AI also brings an increasing number of job opportunities, such as data-based decision engineering and AI technology-added fields. These developments lead to discussions about a growing AI skills gap that might occur as industries adapt to the evolving features of AI technologies. In discussions of the effects of AI on employment issues, the possibility of biased use and deployment of automation technologies might have ethical and moral implications for workforce transition, particularly in the context of Fair AI. These changes can lead to a society where manual labor is undervalued and where risks from retraining the workforce, mainly the unknown investment in learning in an AI-driven economy, make the future of work uncertain. Furthermore, societal changes that lead to irregular working schedules can suppress serendipitous events for those who work irregular hours to upgrade or enhance their formal and informal knowledge. Thus, with a wide range of AI impacts driving employment policies, education and workforce retraining need to be supported by inclusive and comprehensive educational policy measures, which expand collaboration networks and competence models to upgrade and enhance the capacity, accessibility, and opportunity of the workforce. Furthermore, educational measures will open spaces for those who are significantly affected to work and those who will be left out in line with the AI breakthrough. Multi-stakeholders such as industries, administrations, consultants, and educational institutions need to work together to provide the necessary upgrading and upscaling of the workforce. Therefore, human beings must be involved in the process, including creating AI technologies that are implemented and managed. This is consistent with the principles of complementarity, creating and expanding job opportunities through AI development. At the same time, the effects of AI on the workforce and society have the potential to be very harmful and problematic if technological transformations do not take a synergistic approach with appropriate accompanying policies. All countries must prepare policy measures to encourage the growth of AI to increase the positive impact on the workforce [17, 18].

CONCLUSION

The ethical challenges posed by artificial intelligence underscore the need for a balanced approach that combines innovation with societal responsibility. As AI systems increasingly influence decision-making in critical areas, their ethical design and deployment must prioritize transparency, fairness, and privacy. Addressing biases in AI is essential to prevent the perpetuation of social inequities, while robust privacy safeguards are needed to protect individuals' rights. Additionally, the impact of AI on employment requires policies that promote workforce adaptability and inclusiveness. By fostering global dialogue and collaboration across disciplines, society can shape AI technologies that enhance human well-being and support equitable progress. This paper concludes that proactive ethical frameworks are crucial to guiding AI as a force for positive societal change, ensuring AI's alignment with fundamental human values.

REFERENCES

1. Mökander J, Juneja P, Watson DS, Floridi L. The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: what can they learn from each other?. *Minds and Machines*. 2022 Dec;32(4):751-8.
2. Lin Z. Why and how to embrace AI such as ChatGPT in your academic life. *Royal Society open science*. 2023 Aug 23;10(8):230658.
3. Huang C, Zhang Z, Mao B, Yao X. An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence*. 2022 Jul 28;4(4):799-819. ieeexplore.ieee.org
4. Baber H, Nair K, Gupta R, Gurjar K. The beginning of ChatGPT—a systematic and bibliometric review of the literature. *Information and Learning Sciences*. 2024 Jun 5;125(7/8):587-614. [academia.edu](https://www.informalab.it/academia.edu)
5. Longo L, Brcic M, Cabitza F, Choi J, Confalonieri R, Del Ser J, Guidotti R, Hayashi Y, Herrera F, Holzinger A, Jiang R. Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open

- challenges and interdisciplinary research directions. *Information Fusion*. 2024 Jun 1;106:102301. [sciencedirect.com](https://www.sciencedirect.com)
6. Zajko M. Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*. 2022 Mar;16(3):e12962.
 7. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Frontiers in artificial intelligence*. 2023 May 4;6:1169595.
 8. Wallach W, Allen C, Smit I. Machine morality: bottom-up and top-down approaches for modelling human moral faculties. In *Machine ethics and robot ethics* 2020 Sep 10 (pp. 249-266). Routledge.
 9. Kundi B, El Morr C, Gorman R, Dua E. Artificial intelligence and bias: a scoping review. *AI and Society*. 2023 Jan 8:199-215.
 10. Ntoutsis E, Fafalios P, Gadiraju U, Iosifidis V, Nejdil W, Vidal ME, Ruggieri S, Turini F, Papadopoulous S, Krasanakis E, Kompatsiaris I. Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2020 May;10(3):e1356. [wiley.com](https://www.wiley.com)
 11. Nwafor IE. AI ethical bias: a case for AI vigilantism (Allantism) in shaping the regulation of AI. *International Journal of Law and Information Technology*. 2021 Sep 1;29(3):225-40. [academia.edu](https://www.academia.edu)
 12. Ferrara E. Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*. 2023 Dec 26;6(1):3.
 13. Jagatheesaperumal SK, Rahouti M, Ahmad K, Al-Fuqaha A, Guizani M. The duo of artificial intelligence and big data for industry 4.0: Applications, techniques, challenges, and future research directions. *IEEE Internet of Things Journal*. 2021 Dec 31;9(15):12861-85. [\[PDF\]](#)
 14. Li J, Herdem MS, Nathwani J, Wen JZ. Methods and applications for Artificial Intelligence, Big Data, Internet of Things, and Blockchain in smart energy management. *Energy and AI*. 2023 Jan 1;11:100208.
 15. Kuroda Y, Yamamoto G, Kuroda T. Identifying undefined risks: A risk model and a privacy risk identification measure in the privacy impact assessment process. *The Information Society*. 2024 May 26;40(3):202-14.
 16. Idrees R, Maiti A. A Blockchain-Driven Smart Broker for Data Quality Assurance of the Tagged Periodic IoT Data in Publisher-Subscriber Model. *Applied Sciences*. 2024 Jul 5;14(13):5907.
 17. Kumar M, Raut RD, Mangla SK, Ferraris A, Choubey VK. The adoption of artificial intelligence powered workforce management for effective revenue growth of micro, small, and medium scale enterprises (MSMEs). *Production Planning & Control*. 2024 Oct 2;35(13):1639-55. [\[HTML\]](#)
 18. Budhwar P, Chowdhury S, Wood G, Aguinis H, Bamber GJ, Beltran JR, Boselie P, Lee Cooke F, Decker S, DeNisi A, Dey PK. Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*. 2023 Jul;33(3):606-59. [wiley.com](https://www.wiley.com)

CITE AS: Kabiga Chelule Kwemai (2024).The Ethics of Artificial Intelligence in Society. Research Output Journal of Engineering and Scientific Research, 3(3): 1-4.