

ETHICAL CONCERNS IN THE TECHNOLOGY OF ARTIFICIAL INTELLIGENCE: A SOCIOLOGICAL INSIGHT**Aishna Verma**

Research Scholar, Department of Sociology, University of Lucknow, Lucknow

ABSTRACT

One of the most revolutionary technologies of the twenty-first century is artificial intelligence (AI), which has great promise for significant advancements in a wide range of industries, including healthcare, education, banking, transportation, and entertainment. As AI systems become more pervasive in our daily lives, it is crucial to take into account the ethical ramifications of this technological revolution. Therefore, this paper explores the multifaceted relationship between AI and ethics. Based on review of literature and secondary data such as reports, this paper offers sociological insight delving into key ethical considerations, challenges, and the measures needed to ensure for the development of socio-ethical AI.

Keywords:

Artificial intelligence (AI), autonomous systems, ethics, society

INTRODUCTION

Artificial Intelligence (AI) has emerged as one of the most transformative technologies of the 21st century, promising remarkable advances in various fields, from healthcare and finance to transportation and entertainment. These developments may undoubtedly result in large-scale societal changes in the future. As AI systems become increasingly integrated into society, it is imperative to consider the ethical implications that come with this technological revolution. The paper presents the intricate relationship between AI and ethics offering sociological insights, along with important ethical issues, challenges, and recommendations for making AI a constructive force for society.

AI is considered an autonomous technology, and holds immense significance in our increasingly interconnected and complex world. It signifies the culmination of years of research and development in machine learning, neural networks, and robotics. The ability of AI systems to operate autonomously means that, they can perform tasks and make decisions without continuous human guidance. This has profound implications for various fields, such as healthcare, finance, transportation, and more. Therefore, AI as autonomous technology, it's important to clarify what we mean by 'autonomous' in the context of AI. In philosophy, autonomy is defined as 'the capacity of individuals to exercise free will and make autonomous decisions'. On the other hand, as Endsley & Kaber (1999) put forward, "fully autonomous systems carry out their functions on their own, in accordance with predetermined aims".

This implies that they are capable of making decisions without the help of people, even when the context changes (Straube et. al, 2022). For example, self-driving cars use AI algorithms to navigate roads autonomously, making real-time decisions about speed, lane changes, and braking without human intervention. Similarly, AI systems used in industrial automation, such as robots on manufacturing assembly lines, can operate autonomously within predefined parameters. Today, robots are a common sight outside of factories. Robots for everyday assistance and services are one expanding target area of application and will increasingly be used in situations where only humans have traditionally operated, and would assist to improve quality of life (Straube et. al, 2022).

AI and Ethics

Ethical considerations surrounding the use of artificial intelligence (AI) are of utmost importance due to the potential impact of AI systems on individuals, society, and the world at large. This section therefore presents the meaning of ethics in the context of AI technology and the ethical challenges in the AI ecosystem. Aristotle, a Greek philosopher, defined ethics as 'the study of the good', determining the examination of what constitutes a good and fulfilling life for individuals and how one can achieve it through virtuous actions. It is often associated with a broader, more formal, and philosophical approach in determining what is right or wrong. Ethics often

involves the exploration of various moral dilemmas and ethical theories to determine the best course of action. It is often used in professional and academic contexts, such as business ethics, medical ethics, and legal ethics, to establish codes of conduct and standards of behavior within specific fields furthering principles from deontology, utilitarianism, and beyond.

Ethical concern in AI is not only of equal importance as in the other distinguished sciences but an endeavour of utmost priority to be dealt in autonomous technology like artificial intelligence. For example, some of the fundamental ideas that underpin the European approach released by European Commission (2021); AI systems must enable human autonomy and decision-making, be technically sound and risk-averse, ensure that privacy is protected, and be open and transparent. Additionally, they must assure accountability, fairness, non-discrimination, and diversity. Thus, it has become serious area of academic study within AI. Researchers have discovered it is not always possible to define 'fairness' in a way that satisfies all stakeholders (Russell & Norvig, 2021). In order to maintain public faith and trust in the technology, the Global Partnership on Artificial Intelligence, which was established in June 2020, emphasised the importance of developing AI in accordance with democratic principles and human rights (UNESCO, 2021). Also, Leaders of OpenAI released guidelines for 'superintelligence governance' in 2023, anticipating its arrival in less than 10 years (Altman et. al., 2023).

Understanding Ethical Challenges in AI through Sociological Lens

Studying social life in all of its varied manifestations is the goal of sociology. Ethical ideals and ethical endeavours, therefore, are objects of sociological research which contribute to social development (Hoffding, 1905). Ethical considerations surrounding the use of artificial intelligence (AI) are of paramount importance today due to the potential impact of AI systems on individuals, society, and the world at large. Talking of algorithmic bias and fairness, Rose (2023) claims that artificial intelligence and allied technology such as machine learning applications will be biased if they learn from biased data and the developers may not be aware that the bias exists (Grant & Hill, 2023). One well-known example of bias in AI is gender bias in Natural Language Processing (NLP) models. Many AI language models, such as GPT-2 and BERT, have been found to exhibit gender stereotypes and biases in their outputs. For instance, when prompted with sentences like 'He is a doctor, she is a nurse', these models might generate responses like 'He is busy, and she is caring', reinforcing gender stereotypes that associate males with careers like doctoring and females with nurturing roles.

AI systems can inherit biases from their training data, which can lead to discriminatory outcomes. It's crucial to ensure that AI systems are trained on diverse and representative datasets and to regularly audit and mitigate biases in AI algorithms. Other ethical considerations include establishing transparency and accountability for AI-generated decisions, privacy concerns and address potential job displacement and discrimination in the labor market and beyond. As a result of current technical advancements in artificial intelligence and allied technology, new challenges and opportunities relating to the organisation and structure of work are constantly arising.

Dealing with ethical issues surrounding AI that causes prejudice, for instance, obviously tends to assume a prosperous and secure financial future. A study claims that the proportion of worried people in a nation inversely correlates with its probability of poverty and deprivation, i.e., somebody struggling to make ends meet is likely to be more worried about other things than potential AI prejudice. Contrarily, a potential replacement brought on by AI would adversely affect this particular segment of the population more than others in the labour market. Engel & Dahlhaus (2022) also, assert that countries with more structural diversity also tend to have higher proportions of people who are concerned about AI's potential to be used as a tool of discrimination. This may imply that people are more sensitive to the issue of discrimination in societies that are structurally more diverse.

Even in highly automated AI systems, maintaining human oversight and control is crucial, especially in applications like healthcare and law enforcement, where human judgment is critical. One interpretation of modern robotics is that it represents AI. Here, very high standards are required since systems frequently need to react and interact with the outside environment quickly. Currently studies present that, respondents do not trust robotic systems well enough to let them operate in sensitive environments for humans, e.g., in the care domain (Straube et. al, 2022). It depends heavily on whether people trust a technology with which they have had no primary experience. This approach to trustworthiness is particularly effectively represented by ethical principles, such as those that the EU Commission (2021) has released, i.e., the 'human-centric' approach to the creation and application of AI technologies.

Beetz et. al. (2023) put forward, in a situation where people lack first-hand experience with a technology that is still very much in development, relevant drivers of perceived trustworthiness include not only normative beliefs but also attitudes, expectations, psychological needs, and the hopes and fears relating to AI and robots. Drawing

conclusion from the survey, Beetz et. al. (2023) set forth in the context of self-driving cars taking part in road traffic in the future; the mean response of people trusting the technology, rated between “possibly” and “quite probable,” with the middle 50% of responses excluding “probably not” and including “quite probable”.

Therefore, the approach in the development of AI technology necessitates confluence of various disciplines to represent social reality in its true nature. According to Bloomfield (2018), who views science as an amalgam of subjective knowledge rather than objective truths, AI research is influenced by a variety of social and cultural influences. Thus, to comprehend it, one must examine what these conditioning aspects are and how they affect the advancement of AI. Fleck (1987) investigates the reasons why AI research is not a purely scientific process but is instead affected by power struggles within the AI field itself and between AI as a growing field of scientific research and the social and scientific establishments it challenges. In the similar investigations, Forsythe (1993) examines how ‘nontechnical factors’ (Forsythe, 1993) affect AI research, concentrating on how AI systems reflect and embody the cultural values of their creators. She asserts that whether they intend to or not, AI developers incorporate pre-existing values, ideas, and assumptions into their work.

Some sociologists believe that AI is both a socially constituted enterprise and socially constitutive in the sense that it can take social roles, carry out social practises, and establish social relationships when it is used in a social setting. Examining how AI systems ‘penetrate and transform social institutions’ (Schwartz, 1989) and afterwards ‘[redefine] social life’ (Schwartz, 1989) is thus a crucial challenge for sociologists of AI. Others are less convinced by AI’s purported capacity to mimic human behaviour. The social inadequacy of AI systems is a recurring worry. For instance, Collins (2018) addresses the question of ‘Can machines be as intelligent as human beings?’ and argues that while smart machines can demonstrate impressive capabilities to execute mechanical tasks like pattern identification, they lack the ability to ‘see’ and ‘understand’ contexts, a core component of human intelligence. Thus, it is impossible to claim that intelligent machines truly possess human intelligence. Therefore connotes to the idea that a strong social ethical implication is an immediate endeavour to proceed with prior to the successful dissemination of AI in society.

CONCLUSION

Artificial Intelligence as an autonomous technology represents a significant milestone in the advancement of science and technology. Its ability to automate tasks, make complex decisions, and adapt to changing circumstances has far-reaching implications for various industries. However, it faces several challenges that researchers and policymakers must address to harness its potential fully. The ethical considerations, accountability, and the need for robust regulation are essential to ensure AI’s responsible deployment. Ensuring AI is used ethically in areas such as criminal justice, healthcare, and finance is essential. Avoiding misuse for surveillance, discrimination, or manipulation is a key consideration. This further poses need for developing comprehensive regulatory frameworks and ethical guidelines for AI is crucial to strike a balance between innovation and responsible use.

The future of AI lies in successful human-AI collaboration. Designing AI systems that augment human capabilities rather than replace them is a key challenge. Simultaneously, enhancing the transparency and accountability of AI decisions is essential for building trust and accountability. The interaction of AI with humans is among the highest challenges on autonomy of robotic systems, but in any case a clear distinction must be made between the activity (sequence of defined tasks) and the behavior (autonomous decision) of the system. The use of a complex robotic system outside of a controlled environment, such as a production hall, presents challenges that can only be overcome through an interdisciplinary approach and must take into account ethical, legal, and social implications in addition to technical ones. (Straube et. al, 2022)

A technology cannot become widely adopted into society if it is not socially acceptable. For AI applications to be trusted in society, they must be error-free and safe for people. These qualities are undoubtedly required, yet they are insufficient. The same holds true for the usefulness of AI applications, as functionality or usefulness by themselves cannot ensure sufficient social acceptance. As we continue to harness the potential of AI as autonomous technology, striking a balance between innovation and ethics will be paramount in shaping a future where AI benefits society as a whole. Only when ethical standards are upheld simultaneously by the general public and social stakeholders, who ‘trustworthy AI’ must persuade in order to win over enough people, can something become trustworthy.

REFERENCES

- Altman, S., Brockman, G., & Sutskever, I. (2023). Governance of superintelligence. *OpenAI*. Disponible en: <https://bit.ly/3q6NFjv>.
- Beetz, M., Engel, U., Hoyer, N., Kähler, L., Langer, H., Schultheis, H., & Straube, S. (2023). Trustworthiness and Well-Being: The Ethical, Legal, and Social Challenge of Robotic Assistance. *Robots in Care and Everyday Life*, 1.
- Bloomfield, B. P. (Ed.). (2018). *The question of artificial intelligence: Philosophical and sociological perspectives*. Routledge.
- Collins, H. (2018). *Artificial intelligence: against humanity's surrender to computers*. John Wiley & Sons.
- Engel, U., & Dahlhaus, L. (2022). Artificial Intelligence and the Labor Market: Expected Development and Ethical Concerns in the German and European Context. In *Robots in Care and Everyday Life: Future, Ethics, Social Acceptance* (pp. 27-48). Cham: Springer International Publishing.
- Fleck, J. (1987). *Postscript to Development and Establishment in Artificial Intelligence*. Department of Business Studies, University of Edinburgh.
- Forsythe, D. E. (1993). The construction of work in artificial intelligence. *Science, technology, & human values*, 18(4), 460-479.
- Grant, Nico; Hill, Kashmir (22 May 2023). "Google's Photo App Still Can't Find Gorillas. And Neither Can Apple's". *The New York Times*.
- Hoffding, (1905). On the relation between sociology and ethics. *The Sociological Review*, (1), 175-186.
- Rose, Steve (11 July 2023). "AI Utopia or dystopia?". *The Guardian Weekly*. pp. 42-43.
- Russell, Stuart J.; Norvig, Peter. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Hoboken: Pearson. ISBN 978-0134610993. LCCN 20190474.
- Schwartz, R. D. (1989). Artificial intelligence as a sociological phenomenon. *Canadian Journal of Sociology/Cahiers canadiens de sociologie*, 179-202.
- Straube, S., Hoyer, N., Will, N., & Kirchner, F. (2022). The Challenge of Autonomy: What We Can Learn from Research on Robots Designed for Harsh Environments. In *Robots in Care and Everyday Life: Future, Ethics, Social Acceptance* (pp. 57-80). Cham: Springer International Publishing.
- UNESCO Science Report: the Race Against Time for Smarter Development. Paris: UNESCO. 2021. ISBN 978-92-3-100450-6.