



Ethical AI: Necessary Framework for Future

Dr. Sunil Phulre¹, Puneet Nema²

Dept. of CSE, Lakshmi Narain College of Technology, Bhopal, India^{1,2}

Professor¹, Assistant Professor²

sunilp@lnct.ac.in¹, puneetn@lnct.ac.in²

Abstract. *In this modern era of Artificial Intelligence, the advancement of technology is leading to great efficiency and improved quality of life also questioning its use in moral grounds. The use of AI is both beneficial as well as of great harm for our society, country and world depending on its use .with increase in the efficiency and usage of creating thinking of machines there arises some ethical issues such as privacy invasion, discrimination, security risks and many more concerns, creating a great trouble to people's lives and raising the ethical concerns and impacting the social and personal lives of people .The universal application of AI has been deep integrated with the economy and society improving its efficiency and produced benefits along with the extending risk of ethical issues. This following paper will give you an overview about this topic by summarizing the ethical issues which is been created by the AI. this also analyses the challenges faced by peoples in their social lives as well as personal lives because of the increasing ethical issues, also the possible ways to handle it such as discussing the principles and guidelines issued by different organizations, the methods for evaluating ethics of artificial intelligence and the challenges to implement ethics in AI. We hope that this paper provide structured and extensive sketch of AI ethics for fellow researchers in this field.*

Keywords: Artificial intelligence (AI), AI ethics, Ethical issue, Ethical risk, Ethical principle.

Introduction

AI is making a significant impact on the world by fostering innovation, increasing productivity, and tackling global challenges across various sectors. AI is contributing to a wide range of advancements that are making the world more efficient, sustainable, and inclusive. However, it is important to ensure that its development remains responsible and ethical to maximize these benefits. Ethical AI involves designing and implementing artificial intelligence systems in a way that aligns with ethical values, societal norms, and legal requirements. It focuses on ensuring that AI technologies are developed to benefit society, minimize potential harm, and function in a manner that is just, transparent, and responsible.

AI has played a crucial role in advancing various sectors here are given some key point where AI in progressively showing growth towards development- In healthcare, AI has transformed diagnostics, allowing for the early identification of conditions like cancer and heart diseases through advanced imaging and pattern recognition. In education, AI enables customized learning paths for students, helping them learn at their own pace. In finance, AI enhances decision-making with predictive analytics, automating critical functions like fraud detection, risk analysis, and customer service. In transportation, AI enables the development of self-driving vehicles and helps optimize traffic flow, alleviating congestion and increasing safety. In manufacturing, AI-driven automation has improved production



efficiency, quality control, and waste reduction. Predictive maintenance powered by AI ensures minimal downtime, keeping operations running smoothly.

Although AI has significantly advanced several fields, there are a number of issues that need to be resolved. Bias and prejudice are among the primary problems since AI systems have the potential to reinforce biases found in their training data, producing unfair results like racial discrimination in criminal justice systems, as demonstrated by the COMPAS algorithm. Concerns regarding privacy are also raised by the massive data gathering that AI systems necessitate. The Cambridge Analytica affair, in which user data was exploited for political purposes, serves as an example of this. Another significant worry is job displacement, as automation may displace workers in industries like manufacturing and retail. This is demonstrated by the fact that robots are replacing assembly line workers and self-checkout systems are replacing cashiers. Another concern is accountability and transparency, as AI models—such as Amazon's biased recruiting tool—may make judgments that are hard to defend or reverse. Because AI systems are susceptible to cyber-attacks, as demonstrated by the crashes of self-driving cars, security threats are substantial. Artificial intelligence (AI)-generated disinformation, such as deep-fakes, threatens public confidence, while autonomous weapons raise moral concerns about military judgment. Finally, an over-reliance on AI may lead to the loss of critical human abilities, including autonomous cars' ability to make decisions quickly. These issues show how responsible AI development and control are necessary.

1. This article presents a detailed discussion of the ethical issues in AI, covering some of the most significant issues and problems of AI technologies. It also discusses the ethical threats of AI, and suggests the principles and guidelines for the development of responsible AI systems. In addition, the authors describe measures that can be taken to address ethical issues and propose methods for evaluating the ethics of artificial intelligence (AI) systems. This comprehensive review offers key information and serves as a valuable guide to researchers, and practitioners in the field of AI ethics.

2. In this article, a new type of the ethical issues related to AI systems is proposed. This framework is designed to improve the understanding of ethical issues in AI and assists researchers and practitioners in identifying, analyzing, and tackling these concerns. It also discusses the emergence of ethical issues in various stages of the AI system life cycle and proposes a systematic way to tackle those concerns during the development and running of AI systems.

3. The article offers a comprehensive analysis of global AI ethics guidelines, drawing from a review of 146 ethics frameworks published by companies, organizations, and governments across the world. It maps out these principles and provides strategic advice for planning, developing, and implementing AI systems. The paper emphasizes the importance of maintaining ethical guidelines from beginning to end of the AI lifecycle and also presents fundamental principles for minimizing risks and maintaining ethical compliance of AI systems.

4. The paper examines different multidisciplinary approaches to addressing the ethical issues arising within AI, while taking into account ethics, technology, and law. It reviews various methodologies for AI ethics, showing the contributions of multiple fields to the resolution of the challenge at hand that is the nature of AI systems. The effect of this approach is to promote the emergence of comprehensive solutions that not only think in terms of technical issues but also in terms of social, legislative, and ethical issues related to AI.

5. One of the central contributions to this article is its study of techniques for assessing the ethical performance of AI systems. Noting that this is an area that has generally been overlooked in previous



research, it also highlights the importance of AI system testing to establish compliance with ethical guidelines. In describing a range of methods to define AI ethics, the article provides a framework for auditing AI systems to guarantee adherence to ethical standards prior to deployment.

6. The conclusion of the article then focuses on principal challenges in the area of AI ethics, such as the difficulty of defining uniform ethical principles, the need of accountability and the elimination of systemic bias in AI models. It also offers input on future directions for the field, by suggesting significant research issues and relevant fields to explore. This final part is intended to inspire researchers to answer outstanding questions and help to move towards the development of ethical AI.

Methodology

This review covers a wide variety of documents, including academic, organizational, government grey literature sources, and news report. The search of relevant literature was conducted in two phases. In the first phase, the entries or keywords that reflect different terms related to AI ethics are used to search on Google Scholar, Web of Science, IEEE Xplore, ACM Digital Library, Science Direct, Springer Link and Google. The entries or keywords used include: (ethics, ethical, responsibility, responsible, trustworthiness, trustworthy, transparent, explainable, fair, beneficial, robust, safe, private, sustainable) AND/OR (issues, risks, guideline, principle, approach, method, evaluation, assessment, challenge) AND (artificial intelligence, AI, machine learning, ML, intelligent system, intelligent agent). We mainly consider the literature published or released since 2010 and included as many related keywords as possible in titles. In the second phase, we checked the related work of literature found in the first phase, such as the cited articles and other work by the same authors of phase one. As for the ethical AI guidelines, we only collected these documents in English (or with official English translations) and can be visited or downloaded on the internet. A full list with URL links of collected ethical AI guidelines is provided in the Supplementary Materials of this article.

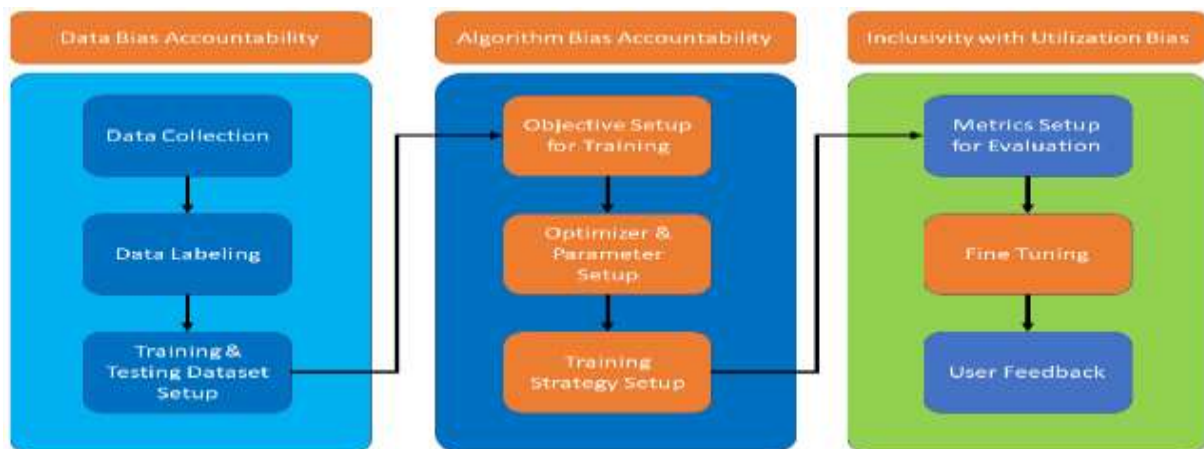


Fig. 1: Bias effects in machine learning.

2.1 Ethical Issues and Risks of AI- In order to cope with the ethical challenges of AI, we must first identify and acknowledge the possible ethical concerns or risks of AI technologies. (Balasubramaniam,



2022) Thereafter, appropriate AI ethical principles, policies, guidelines, and rules (i.e., Ethics of AI) may be eventually crafted. By the proper ethics of AI, we are able to design and implement an ethically behaving AI (i.e., Ethical AI). The term ethical question in artificial intelligence refers to the morally wrong or otherwise problematic consequences of AI (i.e., these problems and risks that are caused by the creation, implementation, and exploitation of AI) that must be dealt with. Many ethical issues, such as lack of transparency, privacy and accountability, bias and discrimination, safety and security problems, the potential for criminal and malicious use, and so on, have been identified from the applications and studies. This part is concerned with the ethical problems and risks of AI.



Fig. 2: Overview Of Artificial Intelligence.

2.2 Overview of Ethical AI

Firstly, four different classes of AI ethical problems in literature are examined in Section and since several of these classes either omit some ethical issues or are too complex to be easily grasped, we presented a new classification to categorize (Changwu Huang, 2022) the AI ethical issues at the individual, societal, and environmental levels. Our novel classification is effortlessly mastered and can include all the present ethical problems and easy to be understood which may assist the understanding and tracing of the ethical problems brought about by AI. Moreover, we try to characterize the ethical concerns that arise as the AI system lifecycle progresses, which could be useful in identifying these problems at the stage of AI system development (Floridi, 2018). The primary purpose of this section is to describe and explain the ethical challenges of AI in order that practitioners be able to identify and appreciate these issues, and subsequently be able to give careful attention to their ongoing investigation of how to deal



with them. The primary contribution to this section is that we have formulated a novel classification scheme of AI ethical problems, which is discussed in an ethical, easy-to-understand format. Furthermore, the ethical problems related to the lifecycle stages of an AI system is presented.

Ethical Guideline And Principles For AI

TRANSPARENCY-AI transparency refers to the explain ability of an agent, i.e., knowing why and how an AI performs a certain action (Balasubramaniam, 2022) The model outputs a particular result for a particular context and, if so, the ability to understand the reasoning used in getting to the result decisions or actions. This aspect is often referred to as "opening the black box of AI. It focuses on making the system's processes clear, explainable, and understandable. Additionally, transparency in AI includes the justification or rationality of the system's design and implementation, as well as the outcomes it generates. Essentially, the design and the choices that the AI system will take will be justifiable and transparent (Paraman, 2023).

FAIRNESS AND JUSTICE: Fairness and justice principle advocates that AI development, deployment and action should follow a set of rules and technologies to avoid potentially harmful, unfair, discriminatory, or unjust outcomes and applications should be carried out in such a way as to avoid discrimination or bias toward, e.g. individuals, communities, or groups. The issue of discrimination and unjust outcomes caused by AI algorithms has emerged as a hot topic in both press and university conversations. As a result, the principles of fairness and justice have attracted much attention in the last years responsibility and Accountability: The principle of responsibility and accountability requires that AI Systems, to guarantee that the designer system, developer system, owner system, and operator system could be auditable, and in this way the designer, developer, owner, and operator of the systems would be answerable actions and decisions of the AI. They are accountable for all damage or adverse outcome that stem from its use. Persons working on the design, development, and application of AI systems are accountable for the Ethical aspects of their operation, such as misuse, and are obliged to advise on these outcomes. Hence, it is also important to develop ways to ensure accountability and accountability for the action of AI systems when being developed and after they have entered deployment.

PRIVACY: The privacy principle (Stahl B. C., 2018) is concerned with maintaining the privacy and data security of these of AI systems. AI systems shall comply with privacy, data protection, and maintain secure data handling practices. This involves establishing strong data governance and management guidelines for all data from and by the AI during its entire life cycle, moreover, data collection, use, and storage should comply with privacy and data protection, moreover data and algorithms should also be secured against theft. In the event of a data breach, AI providers or employers must promptly inform affected individuals—such as employees customers, and partners—to minimize the impact and consequences of the leak.

TRUST: Confidence is critical for AI acceptance by individuals and populations, as it is the bedrock of interpersonal interactions and social functioning. Trust in AI development, deployment, and user relies not just upon the technology itself but upon how the quality of the wider socio-technical system is met thus, trustworthy AI consists not only of the robustness of an AI system, but also latest advances in HX are taking a holistic approach that takes into account the reliability of actors and the procedures as well throughout the system's entire lifecycle.



Future Perspective of Ethical AI

This section discusses future directions for AI ethics research (Torresen, 2018). It argues that, like humans, AI should not rely on a uniform ethical theory, but are able to select a different theory in accordance with the context. This method, termed the "multi-theory method," would allow for AI systems to apply multiple ethical frameworks as appropriate. Additionally, integrating general ethical theories with domain specific ethics, recognized by experts in each field, is important to ensure that AI systems are both ethically sound and accepted by their users. Ethical problems in AI are now being addressed by innovative (Jobin, 2019) machine learning. Ethical principles for ML and AI technologies must be drawn, although it is difficult to coordinate multiple ethical principles during the design of AI, it is critical for the development of good AI ethical AI systems. There is an urgent need for efficient evaluation techniques to measure AI systems, too even though a general evaluation method is still lacking, one may still evaluate the system before deployment. Researchers currently Concentrating on domain-dependent moral judgments, e.g., in the context of autonomous vehicles and healthcare. Furthermore, combining normative ethics (inherent moral capabilities) with evolutionary Ethics (moral development and adjustment) holds potential as a potential route for ethical AI systems.

Conclusion

This study aimed to examine the ethical guidelines organizations have set for developing transparent and explainable AI systems, with a focus on assessing how explain ability can be practically defined. The results indicate that explain ability is closely linked to the transparency and trustworthiness of AI systems. Based on these findings, the study concludes that clearly defining explain ability requirements is a crucial step in creating AI systems that are both transparent and trustworthy. Furthermore, the results highlight that the clarity of an AI system purpose plays a significant role in shaping how its explain ability requirements are defined. This article also explores the complexities and challenges involved in addressing ethical issues in AI and designing AI systems that behave ethically. As AI becomes an increasingly integral part of our society, the success of ethical AI systems will play a critical role in shaping the future. Addressing these ethical issues requires a collaborative effort from AI scientists, engineers, philosophers, users, and policymakers. The article provides a comprehensive overview of AI ethics, summarizing the ethical risks and issues associated with AI, the ethical guidelines and principles proposed by various organizations, strategies for addressing these ethical challenges, and methods for evaluating the ethics (or morality) of AI. In addition to these insights, the article identifies key challenges (Stahl B. C., 2021) in the practical application of AI ethics and outlines potential future research directions in the field (Prem, 2023). Moving forward, we also plan to evaluate the usability and effectiveness of the explain ability components model, with the long-term objective of exploring how explain ability requirements can be effectively integrated into the testing of AI systems

References

[1] H. K. Balasubramaniam, ". Ethical guidelines for solving ethical issues and developing AI systems.," Product-Focused Software Process Improvement: 21st International Conference, PROFES 2020, Turin, Italy, November , vol. 1, no. 21, pp. 331-346, 2022.



-
- [2] Tiwari, Puneet; Phulre Sunil, "Lung Image Segmentation and Compression Using Deep Learning", 2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET), pp. 1-5, 2024, IEEE.
- [3] P. & A. S. Paraman, "Ethical artificial intelligence framework for a good AI society: principles, opportunities and perils," *AI & SOCIETY*, vol. (2), no. 38, pp. 595-611, 2023.
- [4] J. Torresen, "A review of future and ethical perspectives of robotics and AI," *Frontiers in Robotics and A*, vol. 4, p. 75, 2018.
- [5] V. E. Jobin, "The global landscape of AI ethics guidelines," *Nature machine intelligence*, vol. 9, no. 1, pp. 389-399., 2019.
- [6] Nema, Puneet; Phulre Sunil, "A comparative study of smoke and fire detection using deep learning", *Intelligent Computing and Communication Techniques*, pp. 631-636, 2024, CRC Press.
- [7] Stahl, "Ethical issues of AI. Artificial Intelligence for a better future," *An ecosystem perspective on the ethics of AI and emerging digital technologies*, vol. 1, pp. 35-53, 2021.
- [8] Deepak Kumar Rathore, Sunil Phulre, "Machine Learning Based Efficient Prediction of Human Heart Disease by Identifying the Features", *IJIRTM*, vol. 6, no. 6, 2022.
- [9] E. Prem, "From ethical AI frameworks to tools: a review of approaches. AI and Ethics," *AI and Ethics*, vol. 3, no. 3, pp. From ethical AI frameworks to tools: a review of approaches. AI and Ethics, 3(3), 699-716., 2023.
- [10] Nema, Neeraj; Phulre Sunil, "Sentiment analysis and its application on twitter: a survey", *International Journal of Science and Research (IJSR)*, vol. 10, no. 3, pp. 106-109, 2021.
- [11] W. Pu Chen, "AI Fairness in Data Management and Analytics: A Review on Challenges, Methodologies and Applications," in *Methods and Applications of Data Management and Analytics*, Shanghai 201109, China, 202.
- [12] R. Dhaminda B. Abeywickrama, "Model checking human-agent collectives for responsible AI," . In *28th IEEE International Conference on Robot and Human*, pp. 30-39, 2019.
- [13] Verma, Bharti; Phulre Sunil, "Sentiment Analysis based on Naive Bayes and KNN Classifier", *International Journal of Science and Research*, vol. 6, no. 5, pp. 317-319, 2017.
-