Chapter 6

Ethical dimensions in AI innovation

Dr. Shikha Khullar,

Associate Professor, Department of Computer Engineering, Poornima University, Jaipur, Rajasthan

shikha.khullar@poornima.edu.in

&

Sri Satya Sudhir Reddy,

Assistant Professor, Poornima University, Jaipur, Rajasthan

satya.sudhir@poornima.edu.in

Abstract

This chapter aims to discuss specific ethical questions on Prosthetics and AI, and the opportunities and threats indicated by innovation in this field. As various sectors such as healthcare, finance and self-driving systems are impacted by AI technologies, issues like bias, privacy, responsibility and transparency require further discussions. First, the chapter summarizes the historical development of AI ethics, introduces essential frameworks, and discusses industry-specific challenges and future directions of the field and AI in industries, such as Finland's AuroraAI model and AI in the food supply chain. These cases highlight the aptitudes of ethical AI practices and the difficulties of their application. It highlights current measures as well as policy issues and proposes greater intersectoral cooperation on the part of governments, academic institutions, and industry. Moving to the future view, new technologies also bring new ethical issues but can also solve global issues such as global warming and healthcare disparities. Finally, the chapter offers a clear message to stakeholders that the development of AI should not be left to tech-savvy individuals in organizations but must be in line with societal values and sustainability.

Keywords

Ethical AI; AI Frameworks; Bias and Fairness; AI Governance; AI Privacy and Accountability

1. Introduction

AI is one of the greatest inventions of the twenty-first century, covering virtually every industry in today's economy, including healthcare, finance, education, and transportation. AI's capacity to handle, analyse and decide on big data has had profound improvement of efficiency and individualisation (Russell & Norvig, 2020). However, when applied to core operations of an organization, the use of AI has ethical issues such as bias, privacy invasions, and lack of responsibility (Floridi & Cowls, 2022).

These are not limited to concepts of mere professional ethic but they are closely related to social and moral questions. For example, self-driving automobiles or a health system that delivers patient care – these systems have to make decisions that have life and death implications and for this reason they ought to reason ethically just like humans do (Anderson et al., 2005). Likewise, AI's deployment in surveillance and crime forecast arises issues of fairness and right to privacy (Citron & Chesney, 2019). This is especially important because as AI systems become increasingly self-governed, malicious intent and actions also follow the same trend (Binns, 2018).

Dealing with these issues is important for the purpose of achieving the right goals, meaning the goals that reflect society's needs and wants when developing artificial intelligence. Ethical AI is not only a compliance issue but one that falls into proper integration of different principles such as fairness, accountability, and transparency at the AI operations (Floridi, 2019). Current international guidelines for responsible AI include the EU's Ethics Guidelines for Trustworthy Artificial Intelligence and UNESCO's Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021; Smuha, 2019).

The chapter on AI ethics focuses on the growth of AI ethics, current models, concerns of various sectors, and area of future study. It supports the discussion on how to use AI in a responsible manner so that its advantages in various industries will outweigh potential drawbacks.

2. Literature review

Ethical consideration of artificial intelligence has shifted from what was originally mere theoretical considerations to the numerous comprehensive frameworks available today. In 1942 Isaac Asimov outlined three basic rules in Asimov's Three Laws of Robotics which focused on the protection of humans from harm, the 'obstruction' of humans' wishes and the preservation of one's self. Later, Wiener (1988) devoted his work to society consequences of automatons, and Weizenbaum (1976) focused on human values as the exclusive source of moral choices. Such early works laid ground work for deeming AI's ethical issue.

As AI continues to emerge, very appealing methods have been adopted that seek to integrate ethics into practice. Following the principle, Anderson et al. (2005) developed action based ethical theories for use in deciding of machines. Frameworks such as Floridi and Cowls' (2022) five ethical principles that include, and Smuha's (2019) EU Ethics Guidelines for Trustworthy AI also have incorporated some of the principles of accountability and that of transparency. Nevertheless, some issues persist, such as the absence of ways to ensure ethical compliance (Brundage et al.,

2020) and the problem of ethical solution implementation across ecosystems (Stahl, 2022).

AI ethics are thus not simply a matter of general principles but are sensitive to sector-specific considerations. Thus, Chustecki (2024) pointed to biases in diagnosing people and privacy issues, and Hastuti (2023) touched upon such aspects as innovation entrepreneurship and conscience. Recently, Milli (2025) and Caton and Haas (2024) explored dataset failure and misconduct, along with AI algorithm fairness issues. Whereas in autonomous systems, Citron and Chesney (2019) discussed ethical questions related to deepfakes, and Gunning and Aha (2019) provided more focus on interpretability as a result of DARPA's Explainable AI initiative.

However, there are still some gaps, such as the lack of standardized assessment approaches (Stahl, 2022), and guidelines on new technologies like generative AI and quantum computing (Chakraborty, 2025). To overcome these difficulties research should be conducted in close cooperation with specialists from other disciplines and ethical standards should be developed that would be updated constantly.

Table 1: Key References on Ethical Dimensions of AI

Year	Author(s)	Focus/Findings	Research Gaps
1942	Asimov	Introduced Three Laws of Robotics to guide machine behavior.	No practical application in real-world AI systems.
1976	Weizenbaum	Critiqued reliance on computational systems, emphasizing human judgment.	Limited exploration of AI's societal integration.
1988	Wiener	Discussed societal implications of automation in The Human Use of Human Beings.	Did not address emerging AI technologies.
2005	Anderson et al.	Proposed action-based ethical theories for machine ethics.	Lacked focus on large-scale AI applications and systems.
2019	Citron & Chesney	Explored ethical concerns of deepfakes and misinformation.	Limited to media applications without addressing broader AI risks.
2019	Smuha	Outlined the EU's Ethics Guidelines for Trustworthy AI.	Lack of mechanisms for implementation and enforcement.
2020	Brundage et al.	Highlighted mechanisms for trustworthy AI development.	Absence of verifiable standards for ethical compliance.
2021	UNESCO	Published global guidelines on ethical AI.	Lack of regional adaptation and contextual specificity.

2022	Floridi & Cowls	Proposed five ethical principles for AI.	Insufficient focus on practical implementation in diverse sectors.
2022	Stahl	Discussed ethical challenges in innovation ecosystems.	Minimal exploration of sector-specific applications.
2023	Hastuti	Analyzed ethical AI in healthcare, balancing innovation and societal values.	Limited focus on emerging AI technologies like generative AI.
2024	Caton & Haas	Surveyed fairness in machine learning models.	Overlooked specific industries like finance and healthcare.
2024	Chustecki	Reviewed benefits and risks of healthcare AI, focusing on bias and privacy.	Narrow scope excluding global health disparities.
2025	Chakraborty	Addressed emerging ethical issues in generative AI and interdisciplinary approaches.	No detailed framework for addressing challenges in quantum computing.

3. Ethical Frameworks in AI

AI ethical principles are created to regulate AI technologies in an appropriate manner across the society, ethics and often law. These frameworks involve fundamental concepts such as fairness, accountability, and transparency to mention but a few. Combats bias which refers to a scenario whereby an AI system discriminates against individuals or certain groups and inequity where the AI incurs disparate treatment of individuals, and there is a dispute regarding measuring and defining fairness in both Caton and Haas (2024). Legibility concerns involve how decisions are made by AI and who is or can be held legally liable for those decisions at the time of an adverse outcome such as an accident with an autonomous car (Gunning & Aha, 2019). Transparency seeks to make AI systems transparent and explicable, and for this, AI must be capable of scrutiny even if many still remain 'black boxes' (Winfield et al., 2021).

Many documents and recommendations on ethical guidelines within AI were provided by the organizations worldwide. To provide a solution, the EU put forward the Ethics Guidelines for Trustworthy AI which particularly focuses on the principles of human control, non- discrimination and explicability (Smuha, 2019). The United Nations Educational, Scientific and Cultural Organization (UNESCO) has also adopted guidelines referred to as Recommendation on AI Ethics (2021) and the Institute of Electrical and Electronics Engineers' Global Initiative on the ethics of autonomous intelligent systems to ensure that AI is utilized in a manner that respects human rights and societal values. These guidelines are important in setting the standards of ethical Artificial Intelligence around the world (Floridi & Cowls, 2022).

However, there are some challenges with the current ethical frameworks as described by various scholars. Currently there is no international consensus on the key

principles though culture plays a very significant role in the perception and implementation of fairness and accountability as observed by Floridi and Cowls (2022). Additionally, it is critical to note that most frameworks are only guidelines, which reduces their actual enforcement, and, in general, they fail to address many new AI challenges such as generative AI and quantum computing (Chakraborty, 2025). It is also emphasized that clinical implementation necessitates interprofessional cooperation in order to cope with progress in technology.

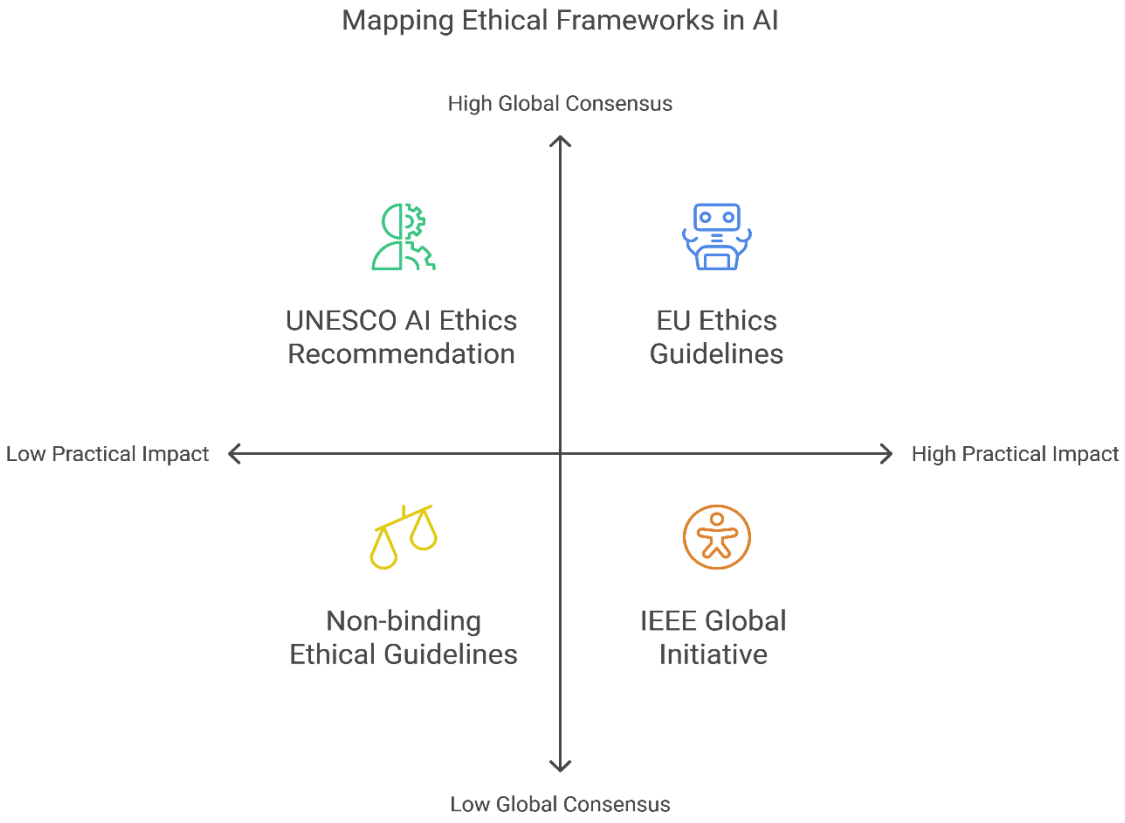

Fig. 1 Conceptual diagram by authors

4. Key Ethical Concerns in AI Innovation

The application of AI technologies has come along with several important ethical issues that have to be resolved in order for a positive impact to be produced. They range from the issues of bias and/ or unfairness, privacy and/ or surveillance, accountability and/ or responsibility, autonomy and control, to the social and/ or economic ramifications of AI systems. These are important concerns to address in order to have AI work fulfilling and effectively respect human values.

Issues of bias and fairness have emerged a key discussion in the last couple of years. AI algorithms are said to carry forward the bias built into the learning data and results

in bias/ unfair decision making in areas like employment, healthcare, and justice systems. For instance, Self-exploitation can lead to reinforcement of racism or sexism for example through biased facial recognition tech or hiring algorithms. Nazer et al. (2023) and Ferrara (2023) describe the sources of bias in training datasets and flawed algorithmic design, and suggest ways to overcome them like using diversified training data and incorporating fairness in algorithms. According to Fu et al. (2020) the use of such algorithms as fairness-enhancing ones can be useful, yet it is still challenging to implement them in such contexts which include this issue of fairness across different groups. Most of these trials are focused on minimizing bias within those algorithms, but they also serve to remind us that the problem of fairness remains a work in progress.

It is even more relevant because of the privacy and surveillance concerns that relate to the use of AI to collect large amounts of data. Technologically advanced systems, for instance, in smart cities or public safety applications pose a threat to individual rights and freedoms by invading their privacy. Thus, the issues of security and privacy that Dhirani et al. (2023) consider to be raising numerous ethical concerns in the context of AI. The General Data Protection Regulation (GDPR) in the region of the European Union offers a legal and conceptual framework allowing for the safeguarding of privacy and the fostering of technological progression. Goel (2024) discusses how AI in cybersecurity, while enhancing security, also poses risks to privacy, necessitating robust frameworks to ensure that privacy rights are upheld while addressing emerging security threats.

It is difficult to designate accountability and responsibility of AI decision-making. The excessively complex models, especially those that utilize machine learning approaches, are hard to explain and therefore can be seen as opaque, or ‘black boxes’. Rai (2020) has pointed out that the absence of openness raises the ethical problem of possible prejudicial outcomes. Proposed solutions like Explainable AI (XAI) which is discussed by Hassija et al., (2024) and Dwivedi et al., (2023) intend to address the interpretability and thus the intelligibility of the AI systems by the stakeholders. In addition, test procedures are recommended, which confirm that AI systems are stable, dependable, and explainable primarily in safety-sensitive industries such as medicine and automotive transportation (Abràmoff et al., 2020).

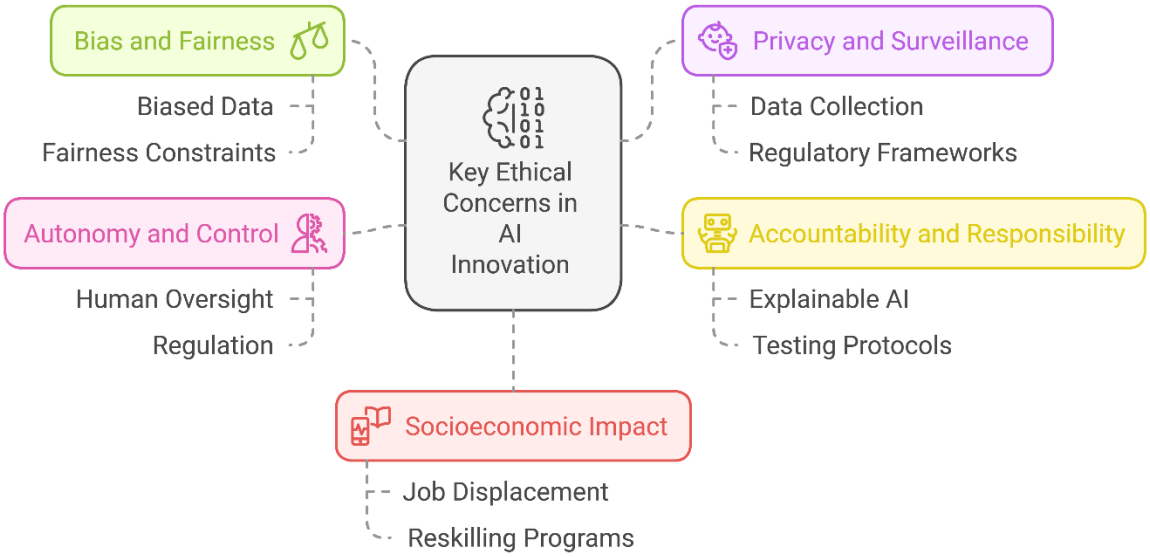

Fig.2 Conceptual diagram by authors

The freedom and management of AI prompt issues over the supervision of individuals besides the likelihood of dependency on AI components. The systems realized in self-driving cars and military drones have certain dangers if decisions made by such systems are incorrect or if the systems themselves stop working. Some of these are cultural issues such as the discretion which must always be retained by the human operator in making the final decision in critical and more so life-threatening situations. Despite the potential of AI to foster autonomous functioning, the technology should be properly managed to prevent negative consequences. Bankins and Formosa (2023) raises the issue of ethics with regards to autonomy of the Artificial Intelligence arguing that measures should be taken so that it always follows the right track or in other words adhere to the Principles of Right Intention as well as self-control shall still lie in the hands of man.

Last but not the least, the centrality of the dimension of impact on the socioeconomic is also increasing due to the emergence of AI technology. It is agreed that AI has the potential to be on the right side of the economic growth; however, it has the potential of causing job loss and economic imbalance. In Bankins and Formosa (2023) writing on the effect of AI on the workforce, the authors argue that to eradicate the vices associated with AI, reskilling programs are inevitable. Chaker and Damak (2024) present the following approaches for the implementation of effective workforce reskilling: Collaboration between academic institutions, governments and players in the sector. Furthermore, Vega et al. (2024) argue that newly created AI technologies have to have an anchoring within the social and cultural realities in order to achieve universal and equal application of AI effectiveness.

These key ethical issues are what, therefore, emphasise the need for a more

comprehensive approach to AI. The problems of bias and privacy, requirements for accountability and autonomy, as well as socioeconomic considerations can be resolved in order to create truly progressive and, at the same time, ethically appropriate AI technologies for all of humanity.

5. Case Studies in Ethical AI Innovation

The case studies from Leikas et al. (2022) and van der Burg et al. (2024) enable the understanding of the ethical issues in AI innovation based on reflecting positive practices and the encountered challenges in the context of AI technologies.

AuroraAI case (Leikas et al., 2022).

The AuroraAI, introduced in Leikas et al. (2022), is one of the biggest projects aimed at implementing AI in the public sector in Finland to enhance citizens' services, especially in the field of medicine and social security. In terms of the ethical principles relate to this project, it emphasis in the fairness, accountability, and transparency of the AI systems. To this end, AuroraAI focuses on increasing inclusiveness of artificial intelligence to reduce prejudice in access to these services. The ethical conflict here pertains to the drive to innovate while at the same time being able to create systems that the citizen can put his trust in especially when it comes to the personal and sensitive area of health. Transparency is a major ethical issue since the initiative seeks to make it easy for citizens to understand how AI systems reach decisions and hold public servants to account. Nonetheless, there are possible problems with ethical governance to be addressed on AuroraAI's way, even as the project premised and concerns its ideals of positivity. These are such things as constant supervision of AI systems to be sure that they are still respecting the ethical values of the society and modification of such systems to feed the new ethical issues.

AI Robots for Agri-Food: Case Study (Van der Burg, et al., 2024).

Consolidated from the AI robots for agri-food case study proposed by van der Burg et al. (2024), this section examines the application of AI and robotics in agriculture and food production, as well as the various important ethical issues arising from such advancements. Another one of the major ethical concerns in this context refers to the autonomy of the artificial intelligence systems, especially when it comes to crop care and the actual act of harvesting. Some questions concern how much direct input from and choices made by people the AI systems should involve when they assume more tasks previously handled by people directly. The developmental consequences of displacement of labour are also significant because the threat of automation encompasses agricultural job for instance in the farming nations. As for this ethical issue, it is important that displaced workers receive access to reskilling resources. Another area of interest is environmental sustainability because adoption of AI in farming techniques should conform to radical and ecological sensitive farming practices which respects the environment. The case study also underscores the issue of accountability to the effect that the robots used in making decisions should have information sufficient to explain its decisions to farmers and consumers. Through discussing these ethical aspects, the study calls for further responsible AI use

in the agri-food system with regard to the various social and environmental impacts.

6. Regulatory and Policy Perspectives

AI ethics is a regulated process, which means that there are multiple attempts of both national and international scales when it comes to setting the standards of ethical AI use. Globally, initiatives and policies were set by the UNESCO and European Union like the EU's Ethics Guidelines for Trustworthy Artificial Intelligence as well as UNESCO's Recommendation on the Ethics of Artificial Intelligence. But they persist in the formulation of sensible and practical standards that can be effectively policed due to continuous technological development, and differing ethical standards globally. In turn, global and regional markets should be governed by a highly integrated system of governments, academia, and the private sector to ensure that the established ethical directions in AI reflect more of the current positive societal and technological trends.

7. Future Directions in Ethical AI Innovation

New AI related technologies opening up opportunities include generative AI, quantum computing, and autonomous system all of which present more ethical dilemmas. These technologies raise issues on privacy, who gets to make a particular decision, and who is held responsible for it. At the same time, AI also has the capability to help in solving the major global challenges within climate change, in health care, and in inequalities. The future of ethical AI innovation focuses on making responsible AI development as the first choice where ethical issues are taken into account in the future. Establishing such culture requires increased awareness of ethical issues in communications among future technologists, creation of synergistic and more adaptive framework and interdisciplinary cooperation to shape up changing ethical outlook of AI.

8. Conclusion

In conclusion, ethical aspect in the development of AI technologies is crucial when aiming at bringing the positive impact of these technologies to society while minimizing drawbacks. Basic concepts of bias, fairness, privacy, accountability, and transparency have turned into crucial matters of discussion in AI. To accommodate this, they have yielded to pressure from international organizations and governments to design ethical frameworks and guidelines but the problem still avails itself in as much as they have not found a way of making them enforceable. New forms of ethical challenges will pour up as AI advances and some of the proposed solutions will need constant observation, flexibility and cooperation with experts of different fields.

Ethical supervision is constantly required, given that AI systems are now involved in the constant regulation of several social capacities. This paper highlights that developers, policymakers, and academics must come up with measures to enhance AI policies and innovations to fit society's objectives and avoid the disadvantages of AI systems. That's why it is critical to promote ethical innovation to guide AI's contribution to solving global issues while protecting human rights and values.

References

1. Abràmoff, M. D., Tobey, D., & Char, D. S. (2020). Lessons learned about autonomous AI: finding a safe, efficacious, and ethical path through the development process. *American journal of ophthalmology*, *214*, 134-142.
2. Anderson, M., Anderson, S., & Armen, C. (2005, November). Towards machine ethics: Implementing two action-based ethical theories. In *Proceedings of the AAAI 2005 fall symposium on machine ethics* (pp. 1-7).
3. Asimov, I. (1942). Runaround. Astounding science fiction, 29(1), 94-103.
4. Bankins, S., & Formosa, P. (2023). The ethical implications of artificial intelligence (AI) for meaningful work. *Journal of Business Ethics*, *185*(4), 725-740.
5. Barnard, P., Marchetti, N., & DaSilva, L. A. (2022). Robust network intrusion detection through explainable artificial intelligence (XAI). *IEEE Networking Letters*, *4*(3), 167-171.
6. Binns, R. (2018, January). Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency* (pp. 149-159). PMLR.
7. Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., ... & Anderljung, M. (2020). Toward trustworthy AI development: mechanisms for supporting verifiable claims. *arXiv preprint arXiv:2004.07213*.
8. Caton, S., & Haas, C. (2024). Fairness in machine learning: A survey. *ACM Computing Surveys*, *56*(7), 1-38.
9. Chaker, B., & Damak, C. (2024). Navigating Tomorrow: Strategies for Effective Workforce Reskilling. In *Reskilling the Workforce for Technological Advancement* (pp. 62-81). IGI Global.
10. Chakraborty, P. P. (2025). Ethical Dimensions of AI Development: Navigating Moral Challenges in Artificial Intelligence Innovation. In *Ethical Dimensions of AI Development* (pp. 59-82). IGI Global.
11. Chustecki, M. (2024). Benefits and risks of AI in health care: Narrative review. *Interactive Journal of Medical Research*, *13*(1), e53616.
12. Citron, D. K., & Chesney, R. (2019). Deepfakes and the new disinformation war. *Foreign Affairs*.
13. Dhirani, L. L., Mukhtiar, N., Chowdhry, B. S., & Newe, T. (2023). Ethical dilemmas and privacy issues in emerging technologies: A review. *Sensors*, *23*(3), 1151.
14. Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., ... & Ranjan, R. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, *55*(9), 1-33.
15. Farooq, S., Farooq, B., Basheer, S., & Walia, S. (2024). Balancing Environmental Sustainability and Privacy Ethical Dilemmas in AI-Enabled Smart Cities. In *Exploring Ethical Dimensions of Environmental Sustainability and Use of AI* (pp. 263-286). IGI Global.

16. Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, *6*(1), 3.
17. Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.
18. Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, 535-545.
19. Fu, R., Huang, Y., & Singh, P. V. (2020). Ai and algorithmic bias: Source, detection, mitigation and implications. *Detection, Mitigation and Implications (July 26, 2020)*.
20. Goel, P. K. (2024). Ethical Considerations in Implementing Artificial Intelligence in Cybersecurity: Balancing Security and Privacy Concerns. In *Redefining Security With Cyber AI* (pp. 72-91). IGI Global.
21. Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI magazine*, *40*(2), 44-58.
22. Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., ... & Hussain, A. (2024). Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*, *16*(1), 45-74.
23. Hastuti, R. (2023). Ethical considerations in the age of artificial intelligence: balancing innovation and social values. *West Science Social and Humanities Studies*, *1*(02), 76-87.
24. Ijaiya, H. Balancing Data Privacy and Technology Advancements: Navigating Ethical Challenges and Shaping Policy Solutions. *Journal homepage: www. ijrpr. com ISSN*, *2582*, 7421.
25. Kuipers, B., McCarthy, J., & Weizenbaum, J. (1976). Computer power and human reason. *ACM SIGART Bulletin*, (58), 4-13.
26. Leikas, J., Johri, A., Latvanen, M., Wessberg, N., & Hahto, A. (2022). Governing ethical AI transformation: a case study of AuroraAI. Frontiers in Artificial Intelligence, 5, 836557.
27. Liao, S. M. (Ed.). (2020). *Ethics of artificial intelligence*. Oxford University Press.
28. Milli, M. Ethical Dimensions of Artificial Intelligence Balancing Innovation and Responsibility. In *AI and Emerging Technologies* (pp. 161-183). CRC Press.
29. Nazer, L. H., Zatarah, R., Waldrip, S., Ke, J. X. C., Moukheiber, M., Khanna, A. K., ... & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health*, *2*(6), e0000278.
30. Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*, 137-141.
31. Roselli, D., Matthews, J., & Talagala, N. (2019, May). Managing bias in AI. In *Companion proceedings of the 2019 world wide web conference* (pp. 539-544).
32. Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Pearson.

33. Smuha, N. A. (2019). The EU approach to ethics guidelines for trustworthy artificial intelligence. *Computer Law Review International*, *20*(4), 97-106.
34. Stahl, B. C. (2022). Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence. *International Journal of Information Management*, *62*, 102441.
35. UNESCO, C. (2021). Recommendation on the ethics of artificial intelligence.
36. Van der Burg, S., Giesbers, E., Bogaardt, M. J., Ouweltjes, W., & Lokhorst, K. (2024). Ethical aspects of AI robots for agri-food; a relational approach based on four case studies. AI & SOCIETY, 39(2), 541-555.
37. Vega, R. P., Rivero, C. S., Castro, A. O., & Bosques, C. V. (2024). Harnessing inclusive innovation to build socially impactful AI: Embracing social impact, cultural diversity, and equity. *Issues in Information Systems*, *25*(4).
38. Wiener, N. (1988). *The human use of human beings: Cybernetics and society* (No. 320). Da capo press.
39. Winfield, A. F., Booth, S., Dennis, L. A., Egawa, T., Hastie, H., Jacobs, N., ... & Watson, E. (2021). IEEE P7001: A proposed standard on transparency. *Frontiers in Robotics and AI*, *8*, 665729.