General Article

The Ethics of Artificial Intelligence: A Critical Examination of Moral Responsibility and Autonomy

Khalid Bashir Hajam[1] & Rachna Purohit[2]

[1]Ph.D. Research Scholars, Department of Education, Indira Gandhi National Tribal University, Amarkantak, M.P.

[2]Ph.D. Research Scholars, Department of Education, Indira Gandhi National Tribal University, Amarkantak, M.P.

Email: rachnapurohit1993@gmail.com

Abstract

The present article aims to critically examine the ethical considerations surrounding artificial intelligence (AI) with respect to moral responsibility and autonomy. The authors illustrate the development of AI requiring careful recognition of its potential impact on society and the demand for considerable accountability and transparency in the design and utilization of AI systems for addressing the ethical issues persisting regarding the implementation of AI. The paper further discusses the crucial role of AI ethical frameworks and guidelines built on moral obligations and autonomy for contemporary society and the necessity of collaboration between different stakeholders to ensure the safe use of AI that promotes the well-being of the people. The article also mentions the various initiatives of the Indian government taken to develop and implement ethical AI in the country and specifies the indispensable task of NEP 2020 in promoting AI ethically in the Indian education system. Moreover, the current study suggests the prerequisite research and dialogue on this salient study area of AI to ensure its ethical development and application for the social security and digital safety of the individual as well as the community as a whole.

Keywords: Artificial Intelligence (AI), Ethics, Moral Responsibility, Autonomy, Transparency.

Overview

Artificial Intelligence (AI) is a common expression that pertains to the making of computer-based systems efficient in executing activities that specifically necessitate the intelligence of humans, including acquiring knowledge, resolving problems, and making decisions. Russell and Norvig (2010) defined AI as "the study of agents that receive percepts from the environment and take actions that affect that environment". They manifested AI as a sub-discipline of computer science that centres making intelligent agents capable of reasoning, perceiving, and acting autonomously. Domingos (2015) provides a more general definition of AI, describing it as "the quest to build machines that can think like humans", whereas, Goodfellow et al. (2016) expounded it as "the ability of machines to perform tasks that would normally require human intelligence, such as visual perception, speech recognition, decision-making, and language translation". Therefore, AI can be described as the exploration of developing intelligent agents that have the ability to autonomously reason, perceive, and take action.

Artificial intelligence has its period of development from the early days

of computer science. The concept of machines that can simulate human intelligence has fascinated researchers and scientists for decades, leading to the development of various approaches and techniques for creating AI systems. The earliest developments in AI can be dated back to the 1950s, when researchers began exploring the possibility of creating machines that could "think" like humans (Russell & Norvig, 2010). In the beginning, AI systems were created with the intention of carrying out elementary tasks like playing games and resolving predicaments related to mathematics. These mentioned systems were founded on rule-based algorithms that were programmed to follow a set of predefined rules. In the 1960s and 1970s, researchers began exploring new approaches to AI, including machine learning and neural networks (Jordan & Mitchell, 2015). The 1980s and 1990s saw significant advancements in AI, including the development of expert systems, which were designed to mimic the decision-making abilities of humans (Russell & Norvig, 2010). These mechanisms were utilized in a vast array of implementations, encompassing everything from medical diagnosis to economic prediction. In the early 2000s, researchers began exploring new approaches to AI, including deep learning and natural language processing (Floridi & Cowls, 2019). Today, artificial neural networks are utilized by deep learning algorithms for learning and improving their performance over time, making them exceptionally skilled in tasks like image and speech recognition. Processing of Natural language algorithms is designed to understand and interpret human language, enabling machines to interact with humans more naturally and intuitively. Today, AI is a rapidly growing field with applications in a wide range of industries, including healthcare, finance, and transportation, which has been driven by advancements in computing power and data analytics, as well as increased interest and investment from industry and the government (Brynjolfsson & McAfee, 2014)

The advent of AI has been influencing almost every sphere of life. Experts have been concerned about the drastic impact of AI on individuals and society regarding its various issues, including ethical concerns. The realm of ethics in artificial intelligence (AI) is multifaceted and swiftly evolving, as its creation and implementation bring up a broad spectrum of ethical dilemmas. The foremost ethical concern in AI pertains to partiality and inequity. It can learn from data, and if that data contains biases, the resulting AI systems can perpetuate and even amplify those biases (Crawford & Calo, 2016). For instance, AI systems employed in recruitment or lending assessments may exhibit discrimination against particular groups based on attributes like race, gender, or age. Additionally, transparency and accountability are also ethical concerns that are significant in AI. AI systems are frequently complicated and obscure, which can make it challenging to comprehend how they reach their decisions. This lack of transparency can make it difficult to identify and correct biases or errors in it (Floridi & Cowls, 2019). Additionally, it can be challenging to allocate the responsibility for the actions of AI systems, particularly in cases where the system's decision-making processes are not transparent. In addition, another ethical issue concerning AI is how it affects human autonomy and decision-making. As these systems become more sophisticated, they might be utilized to make choices that can have substantial consequences on individuals' lives, such as decisions regarding healthcare or criminal justice. This raises questions about the extent to which humans should be involved in these decisions, and whether AI systems can be trusted

to make fair and ethical decisions (Selbst et al., 2019). A variety of ethical frameworks and principles have been proposed to address these and other ethical concerns in AI. These include approaches based on principles such as transparency, accountability, and fairness (Floridi & Cowls, 2019), as well as principles such as beneficence, non-maleficence, and respect for autonomy (Taddeo & Floridi, 2018) while some researchers have proposed the development of ethical codes of conduct and regulations for the development and deployment of AI systems (Bostrom & Yudkowsky, 2014). Thus, ethics in artificial intelligence is a complicated area of study with a wide range of ethical problems and concerns including issues of bias and discrimination, transparency and accountability, and the impact of AI on human freedom for making decisions.

Moral responsibility in the context of Artificial Intelligence

Moral responsibility is a central concept in discussions of ethics and artificial intelligence (AI). As AI systems gradually evolve and become independent, they are able to make decisions that can greatly affect people's lives. This raises concerns about who should be held accountable for the actions of these systems and how accountability should be assigned. One of the challenges in assigning moral responsibility in the context of AI is the fact that these systems are often designed to learn and make decisions on their own, without direct human intervention. This raises questions about whether responsibility for the actions of an AI system should rest with the humans who created or deployed the system, or with the system itself (Dastani & Yazdanpanah, 2022). Additionally, there may be multiple factors involved in the development and deployment of AI, further complicating questions of responsibility. Another challenge is the fact that these systems can be opaque and hard to understand. This can make it difficult to identify and correct biases or errors in the system's decision-making processes, and can also make it strenuous to assign responsibility for the actions of the system (Floridi & Taddeo, 2016). To deal with these challenges, some researchers have suggested the development of new ethical frameworks and principles regarding moral responsibility with respect to AI. One approach is to consider the degree of control that humans have over an AI system, with greater control implying greater moral responsibility (Gianni et al., 2022). Others have proposed the development of new legal frameworks that would clarify the allocation of responsibility for the actions of AI systems. Apart from these technical and legal remedies, certain experts have contended that a more extensive cultural transformation may be required to guarantee that AI systems are created and utilized in a moral and reliable way. This could involve promoting a greater awareness of the ethical implications of AI, as well as encouraging greater collaboration between experts in AI and ethics (Floridi & Taddeo, 2016).

Artificial intelligence Autonomy in the context of AI

Artificial intelligence (AI) autonomy means the ability of an AI system to make decisions and take actions without direct human input or supervision. The level of autonomy in these systems can vary widely, ranging from simple rule-based systems that make decisions based on predefined criteria, to more sophisticated machine-learning systems that can adapt and learn from experience. As AI systems become more autonomous, it becomes increasingly important to ensure that they act in ways that are consistent with ethical and legal norms and that those responsible

for these systems are held accountable for their actions (Floridi, 2019). In order to address these challenges, some researchers have proposed the use of "explainable" AI, which allows users to understand how AI systems arrive at their decisions and makes it easier to identify and correct errors or biases (Doshi-Velez & Kim, 2017).

Moral Responsibility and AI Autonomy for Ethical AI

The relationship between autonomy and moral responsibility in connection with artificial intelligence (AI) is intricate and has many aspects. As AI systems gain more independence, they can make decisions and execute actions in the absence of human intervention or supervision. This growing autonomy brings up critical ethical inquiries about the moral responsibility of those engaged in their creation and implementation. One way to think about the relationship between autonomy and moral responsibility is through the concept of "explainability". As AI systems grow more autonomous, they also become more difficult to understand and explain. This can make it challenging to determine who is responsible when an AI system makes a decision that has negative consequences (Bostrom & Yudkowsky, 2014). To address these problems, some researchers have proposed the development of new ethical and legal frameworks that are better suited to the unique challenges posed by autonomous AI systems (Bryson, 2018). Others have suggested the use of explainable AI, which allows users to understand how AI systems arrive at their decisions and makes it easier to identify and correct errors or biases (Doshi-Velez & Kim, 2017). Besides these technical solutions, it is also crucial to contemplate the societal and organizational factors that can influence the moral responsibility of those involved in AI design and deployment. For example, the incentives and pressures faced by developers and organizations may influence their willingness to prioritize ethical considerations (Floridi, 2019). Thus, it can be illustrated that autonomy and moral responsibility with respect to artificial intelligence are intertwined concepts. As AI systems grow more self-reliant, it becomes progressively crucial to guarantee that they behave in a manner that aligns with ethical and lawful standards and that those accountable for these systems are held answerable for their conduct.

Implications of Autonomy and Morality in AI

The artificial intelligence (AI) autonomy for moral development is significant and has important implications for society as a whole. With the increasing independence of AI systems, they can now make decisions that carry significant ethical implications. This raises crucial inquiries regarding the part AI plays in shaping ethical progress and the effects of these systems on both individuals and the community. One potential implication of AI autonomy for moral development is the potential for these systems to shape our moral beliefs and values. As these systems become more ubiquitous, they may become a major source of moral guidance for individuals and society. This raises important concerns about who is responsible for the development and programming of these systems, and how we can ensure that they reflect our shared values and beliefs (Bostrom & Yudkowsky, 2014). Another potential implication of AI autonomy for moral development is the potential for these systems to exacerbate existing moral biases and inequalities. AI systems are only as impartial and unbiased as the data they are trained on, and if this data reflects existing societal biases and inequalities, then the decisions made

by these systems will also reflect these biases (Crawford et al., 2018) giving rise to concernments about how we can ensure that AI systems are designed and trained in a way that is fair and unbiased. In addition to these issues, there are also important questions about how AI autonomy will impact our capacity for moral reasoning and decision-making. As AI systems become more autonomous, they may become a crutch for our own moral reasoning, leading to a decline in our own capacity for moral judgment and decision-making (Gunkel, 2018). Therefore, the implications of AI autonomy for moral development are significant and important for society as a whole. As we progress in the development and implementation of AI systems, it is crucial to take into account their implications and adopt measures to ensure that they align with our collective principles and attitudes. They should be created and trained impartially and equitably and should not compromise our ability to exercise moral reasoning and decision-making.

Transparency and Explainability: Road to Ethical AI

Transparency and explainability play an important role in promoting ethical artificial intelligence (AI) by increasing accountability, trust, and understanding of AI systems. Transparency implies having the capability to obtain data about the internal operations of AI systems, which encompasses the data utilized to instruct the model, the algorithms employed, and the process of decision-making. It is essential for accountability and trust in AI systems, particularly in sensitive areas such as healthcare, finance, and criminal justice (Burrell, 2016). Lack of transparency can also lead to biases and discrimination in AI systems, as it can be difficult to identify and correct errors or biases in the system (Mittelstadt et al., 2019). Explainability is another crucial factor to create ethical AI systems which pertains to the capability of comprehending the process and rationale behind the decisions or suggestions made by AI systems. This is specifically important in instances where the consequences of the decision are significant. Explainability allows for increased accountability and trust in AI systems, as it allows stakeholders to understand the rationale behind the decision (Lipton, 2018). Lack of explainability can also lead to a lack of trust in AI systems, particularly if they operate as "black boxes" with no clear explanation of how they arrived at their decision (Mittelstadt et al., 2019). Hence, several approaches have been proposed to increase transparency and explainability in these systems. One approach is to develop explainable AI (XAI) systems, which are designed to be transparent and provide explanations for their decisions (Adadi & Berrada, 2018). This can be achieved through techniques such as rule-based systems, decision trees, or visualizations of the decision-making process. Another approach is to develop post-hoc explainability techniques, which provide elucidation of the decisions made by the black-box system of AI. This can be achieved through techniques such as feature importance analysis or perturbation analysis (Ribeiro, Singh, &Guestrin, 2016). Thus, transparency and explainability are important for promoting ethical AI by increasing accountability, trust, and understanding of AI systems through approaches such as XAI systems or post-hoc explainability techniques which can help to enhance transparency and explainability in AI systems.

Initiatives for Developing Ethical AI in India

The government of India has proposed manifold initiatives for developing and implementing ethical AI in the country. These initiatives aim to ensure that AI systems are outlined and used in

ways that promote human welfare, respect human dignity, and are guided by fairness, justice, and equality principles. Some of the key initiatives and frameworks proposed by the government are as follows:

- **National Strategy for Artificial Intelligence (NSAI):** The NSAI was launched in 2018 to mentor the development and implementation of AI in the country. The strategy includes a focus on developing ethical AI systems that are transparent, accountable, and fair.
- **National Program on AI:** The National Program on AI was launched in 2020 to promote the development, growth, and deployment of AI in India. The program includes a focus on developing ethical AI systems that align with national values and ethics.
- **Responsible AI for all**: Niti Aayog, the planning commission of the Government of India, has published Responsible AI approach documents in collaboration with the World Economic Forum Centre for the Fourth Industrial Revolution. These documents aim to establish elaborated ethics doctrines for the design, development, and implementation of AI in India.
- **Standards Setting Bodies:** Various standards setting bodies, such as the IEEE Global Initiative on Ethics and Autonomous and Intelligent Systems and Ethically Aligned Design, have developed ethical frameworks for guiding the development and use of AI.
- **AI FOR ALL - Approach Document for India:** Niti Aayog has also released an approach document for responsible AI, which provides comprehensive AI ethics principles to mentor the overall planning, development, and execution of AI in the country.
- **FAT/FATE Principles:** AI regulations in India are designed to include FAT (fairness, accountability, transparency) or FATE (fairness, accountability, transparency, and ethics) principles to assure accountable, ethical, safe, and responsible implementation of AI tools.
- **National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS):** The launch of NM-ICPS in 2018 opened the doors to promote research and development in interdisciplinary areas such as robotics, AI, and automation, focussing on developing ethical AI systems that are socially accountable as well as sustainable.
- **Data Protection Bill:** The Data Protection Bill, which is currently under consideration, includes provisions for protecting the privacy and personal data of individuals in the development and use of AI systems (NitiAyog, 2018).

These initiatives and frameworks aim to ensure that AI technologies are established and used in a manner that coordinates with ethical principles, respects human rights, and fosters the welfare of the society. By incorporating ethical considerations into AI development and deployment, India is taking steps towards responsible and ethical AI use responsibly as well as sustainably.

NEP 2020 and Ethical AI

NEP2020 is a comprehensive policy framework for education in India. While it primarily focuses on transforming the education system, it also recognizes the importance of integrating technology into education. As a component of this incorporation, the NEP 2020 accentuates the significance of ethical considerations when creating and employing AI. Ethical considerations

with regard to AI simply ensure that the construction and operation of AI systems are directed by doctrines such as openness, liability, impartiality, and confidentiality. The NEP 2020 recognizes these principles and states that the use of AI in education must be based on ethical considerations. For this, the National Education Policy (NEP) 2020 recommends several measures regarding ethical AI in education. Here are some of the key recommendations:

- **Incorporating ethical AI principles in the curriculum:** The NEP 2020 proposes that the curriculum should include ethical deliberations during the creation and use of AI. This will guarantee that students have the required expertise and knowledge to build and utilize AI systems ethically and responsibly.
- **Encouraging research on ethical AI:** The NEP 2020 recommends that research on ethical AI be encouraged and supported. This will facilitate the creation of AI systems that possess fairness, transparency, and accountability.
- **Developing guidelines for AI in education:** The NEP 2020 recommends that guidelines for developing and using AI in education be developed. These guidelines should incorporate ethical considerations and ensure that AI systems are responsibly and sustainably used and build.
- **Ensuring transparency and accountability in AI systems:** The NEP 2020 recommends that AI systems used in education be transparent in making decisions and using data. This will ensure accountability and fairness in developing and using AI systems.
- **Developing a framework for privacy protection:** The NEP 2020 recommends the development of a framework for protecting the privacy of students and other stakeholders in developing and using AI systems in education (India, Ministry of Education, 2020).

Therefore, NEP 2020 acknowledges the significance of ethical AI in education and proposes suggestions to guarantee the responsible and sustainable creation and utilization of AI systems.

Benefits of AI in Education

The emergence of Artificial Intelligence in Education (AIEd) has completely revolutionized every aspect of the educational system. According to Baker and Smith (2019) the AI technologies implemented in educational settings can be categorized as learner-facing, teacher-facing, and system-facing AIEd approaches. These approaches have immense advantages that have been discussed below:

- **Learner-facing AIEd approach:** It uses AI tools such as Intelligent Tutoring Systems (ITS) that are beneficial for students in learning the subject matter. These adaptive learning systems simulate one-to-one personalized tutoring (such as MOOCs) which provides an edge for making decisions regarding the learning path of every individual learner, providing cognitive scaffolding, and engaging students in dialogue. These systems have extensive potential to promote distance education via the implementation of modules in the instructional process, where one-to-one human interaction is negligible. Based on learner models, ITS fosters an online collaborative and interactive learning environment. Virtual reality, a form of ITS, can be helpful in engaging and guiding students in authentic virtual reality and game-based learning environments. The real-time feedback mechanism of AI

applications provides learners with guidance and prompts when they get confused and stuck during their learning process.

- **Teacher-facing AIEd approach:** It uses AI tools that are helpful in supporting and reducing the workload of teachers regarding attendance, administration, assessment, feedback, and detection of plagiarism via automation. Studies have proven that AI systems such as Intelligent Agents save time for teachers teaching online by leaving the most repetitive tasks on the system. This, in turn, encourages teachers to focus on more creative works. AI tools can also help in providing teacher feedback without intervening in the privacy of the feedback provider, thus providing insights to teachers into their teaching methods and strategies. The adaptive systems help in extracting information regarding the performance of students that can help teachers perform diagnostic tasks for presenting proactive guidance to the students in need.
- **System-facing AIEd approach:** It uses AI tools that are used by managers and administrators at the institutional level. AI applications can be helpful in simplifying the work and managing the time of the administrative staff by accurately predicting the admission decisions, student retention and drop-out, maintain e-portfolios, perform automated grading, keeping a record of students' credentials etc.

Conclusion

The significance of ethics in the realm of AI has grown significantly in recent times due to the widespread use of AI in divergent fields. Thus, it is crucial to scrutinize the moral responsibility and independence of those involved in creating and deploying AI systems. The article provides a thought-provoking analysis of the ethical considerations surrounding the development and utilization of AI by exploring the issues of moral responsibility and autonomy in the context of AI and examining the challenges and opportunities that arise with the advancement of this technology. The authors argue that the development of AI requires careful consideration of its potential impact on society and that there is a requirement for considerable transparency and accountability in the design and use of AI systems. The authors suggest that this can be acquired with the development of ethical frameworks and directives that prioritize the principles of fairness, accountability, and transparency. Furthermore, the article highlights the need for greater collaboration between varied stakeholders, including legislators, industrial leaders, and the public, to ensure that the development of AI is mentored by ethical contemplations that prefer the interests of society as a whole. Finally, the researchers underscore the need for continued research and dialogue on this important topic to ensure that AI is made to evolve and used in a form that promotes the greater good of humanity.

References

Adadi, A., &Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/access.2018.2870052

Baker, T., & Smith, L. (2019). *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges.*Nesta Foundation. Retrieved August 14, 2023, from https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf

Bostrom, N., &Yudkowsky, E. (2014). The ethics of artificial intelligence. In *Cambridge University Press eBooks* (pp. 316–334). https://doi.org/10.1017/cbo9781139046855.020

Brynjolfsson, E., & McAfee, A. (2015). The second machine age: work, progress, and prosperity in a time of brilliant technologies. *Choice Reviews Online*, *52*(06), 52–3201. https://doi.org/10.5860/choice.184834

Bryson, J. J., Diamantis, M., & Grant, T. D. (2017). Of, for, and by the people: the legal lacuna of synthetic persons. *Artificial Intelligence and Law*, *25*(3), 273–291. https://doi.org/10.1007/s10506-017-9214-9

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 205395171562251. https://doi.org/10.1177/2053951715622512

Crawford, K., &Calo, R. (2016). There is a blind spot in AI research. *Nature*, *538*(7625), 311–313. https://doi.org/10.1038/538311a

Dastani, M., &Yazdanpanah, V. (2022). Responsibility of AI systems. *AI & Society*, *38*(2), 843–852. https://doi.org/10.1007/s00146-022-01481-4

Domingos, P. (2016). The master algorithm: how the quest for the ultimate learning machine will remake our world. *Choice Reviews Online*, *53*(07), 53–3100. https://doi.org/10.5860/choice.194685

Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1702.08608

Floridi, L. (2018). Soft ethics and the governance of the digital. *Philosophy & Technology*, *31*(1), 1–8. https://doi.org/10.1007/s13347-018-0303-9

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A*, *374*(2083), 20160360. https://doi.org/10.1098/rsta.2016.0360

Gianni, R., Lehtinen, S., &Nieminen, M. (2022). Governance of responsible AI: From ethical guidelines to cooperative policies. *Frontiers in Computer Science*, *4*. https://doi.org/10.3389/fcomp.2022.873437

Goodfellow, I., Bengio, Y., &Courville, A. (2016). Deep learning. In *MIT Press eBooks*. https://dl.acm.org/citation.cfm?id=3086952

Gunkel, D. J. (2017). The other question: can and should robots have rights? *Ethics and Information Technology*, *20*(2), 87–99. https://doi.org/10.1007/s10676-017-9442-4

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Joshi, K. (2022, October 14). *AI Regulations & Laws In India: A Step Towards Ethical AI Use*. XAI. Retrieved August 14, 2023, from https://xai.arya.ai/article/ai-regulations-laws-in-india-a-step-towards-ethical-ai-use#:~:text=The%20planning%20commission%20(Niti%20Aayog,sectoral%20regulatory%20guidelines%20encompassing%20privacy%2F

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, *61*(10), 36–43. https://doi.org/10.1145/3233231

National Education Policy. (2020). Retrieved August 13, 2023, from https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf

NitiAayog. (2018). *National Strategy for Artificial Intelligence*. Retrieved August 13, 2023, from https://niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf

Rakesh, D. (2023, February 7). *Ethics for AI | Challenges & Resolution. INDIA Ai*. Retrieved August 14, 2014, from https://indiaai.gov.in/article/ethics-for-ai-challenges-resolution

Ribeiro, M. T., Singh, S., &Guestrin, C. (2016). "Why Should I Trust You?" *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939778

Roy, A. (2022, September 13). *Why India needs to talk more about AI Ethics. INDIA* Ai. Retrieved August 14, 2023, from https://indiaai.gov.in/article/why-india-needs-to-talk-more-about-ai-ethics

Russell, S. J., &Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3rd ed.). Prentice Hall.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., &Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3287560.3287598

Taddeo, M., &Floridi, L. (2018). How AI can be a force for good. *Science*, *361*(6404), 751–752. https://doi.org/10.1126/science.aat5991

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., ...& Schwartz, O. (2018). *AI now report 2018* (pp. 1-62). New York: AI Now Institute at New York University