ETHICS IN AI: WORLDWIDE IMPACTS AND EVOLVING TRENDS

Binoy Checkonthayil♦

Dharmaram College, Bengaluru

Abstract

Artificial Intelligence (AI) is revolutionizing various aspects of society and the burgeoning integration of AI systems into daily life has exacerbated the ethical implications of their deployment worldwide. AI ethics encompasses a wide range of issues, including privacy, bias, accountability, transparency, and the societal consequences of automation. The creation of thorough ethical rules has lagged behind the quick growth of AI technology, creating difficulties in guaranteeing the responsible design and application of AI systems. Because AI systems frequently demand enormous datasets, which may expose sensitive personal information, privacy concerns are raised. The potential of artificial intelligence to deduce facts that people might not have voluntarily disclosed further complicates this problem. AI bias is yet another serious ethical issue as biases that already exist in the data that AI systems are trained on have the potential to be reinforced by these systems which leads to unfair treatment and discrimination, particularly against marginalized groups. Thus, accountability in AI which is essential for addressing the ethical concerns are needed in establishing legal frameworks to supervise the deployment of AI. In AI

♦ **Binoy Checkonthayil** is an Indian Catholic priest from Kerala who belongs to the religious congregation Carmelites of Mary Immaculate (CMI). He holds an M.Phil. Degree from Christ University, Bengaluru and Ph.D in Psychology from Nagpur University. He also has an M.A. and an M. Ed. in Sociology. He teaches Psychology in Dharmaram Vidya Kshetram and is a visiting Professor of Darsana Institute of Philosophy, Wardha. In Ohio, USA, he worked as a teaching assistant. He spent six years as the principal of the BJM Carmel Academy, CBSE Senior Higher Secondary School in Chandrapur. He is the author of the book 'Mentally Retarded Children: Stress and Coping Resources Among Indian Mothers' published by Scholars Press, London. He is an established motivational speaker. Email: binoycmi@gmail.com

ethics, transparency is equally essential. Globally, the approach to AI ethics varies significantly across different regions. Thus, the paper examines the global impacts and evolving trends in AI ethics, exploring the balance between technological advancement and moral responsibility, as well as the role of international cooperation in addressing AI ethics. The establishment of global standards and agreements can harmonize ethical practices and ensure that AI benefits are distributed equitably. Thus, the ethical implications of AI are complex and multifaceted, requiring a coordinated effort from governments, industry and civil society. Measures for the use of AI's advantages while reducing its risks by tackling the issues of privacy, bias, accountability, transparency and societal effect are the need of the hour.

Keywords: Artificial Intelligence (AI), Ethics, Global Impacts, Global Trends, transparency

Introduction

Artificial Intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions - with some degree of autonomy - to achieve specific goals[1]). Both software and hardware systems incorporate Artificial intelligence (AI. Software-based systems can function in a virtual environment and include voice assistants, image analysis software, search engines, speech and face recognition systems and advanced robots, autonomous cars, drones and Internet of Things applications.[2] Artificial Intelligence (AI) has become a cornerstone of modern technological advancement, permeating various aspects of our daily lives. AI is now present in many facets of daily life, from entertainment and transportation to healthcare and education. Artificial Intelligence presents both exceptional opportunities and challenges as it develops and becomes more integrated into society frameworks.[3] The ethical ramifications of
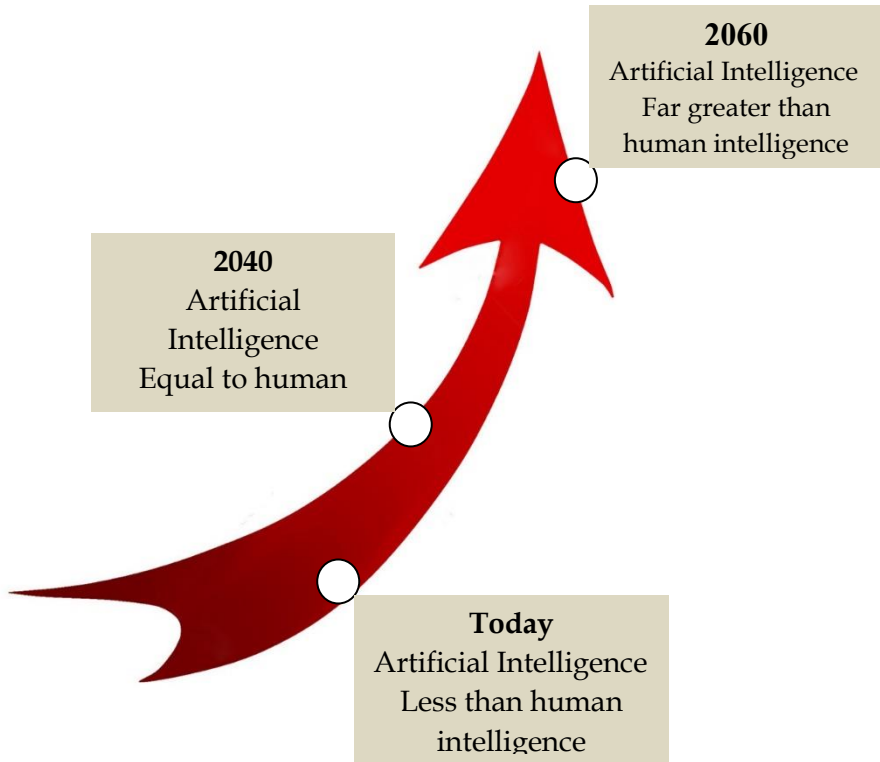
---

[1] Sheikh, H., Prins, C. and Schrijvers, E., *Artificial Intelligence: Definition and Background*. In Mission AI: The New System Technology. Sheikh, H., Prins, C. and Schrijvers, E., (Eds.). Research for Policy. Springer, Cham. https://doi.org/10.1007/ 978-3-031-21448-6_2, 2023.

[2] Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E. and Winfield, A., The ethics of artificial intelligence: Issues and initiatives. Study Panel for the Future of Science and Technology, EPRS - European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_ST U(2020)634452_EN.pdf, 2020.

[3] Banaeian Far, S. and Imani Rad, A., Internet of Artificial Intelligence (IoAI): the emergence of an autonomous, generative, and fully human-disconnected community. Discov. Appl. Sci. 6:91 (2024).

AI development and application are among the most urgent issues it raises. Fig. 1 illustrates the development and future of AI over the years.

2060
Artificial Intelligence
Far greater than
human intelligence

2040
Artificial
Intelligence
Equal to human

Today
Artificial Intelligence
Less than human
intelligence

Fig 1: Rapid growth and future of AI

Maintaining public confidence and promoting fair results require that AI systems be developed, deployed, and used in an ethically sound manner.[4] AI ethics cover a wide range of topics, such as accountability and responsibility, privacy and security, transparency and explainability, autonomy and control, and bias and fairness. In addition to being technical difficulties, these ethical considerations have a strong cultural, legal and human rights foundation. The

---

[4] Ortega-Bolaños, R., Bernal-Salcedo, J., Germán Ortiz, M. Sarmiento, J.G., Ruz, G.A. and Tabares-Soto, R., Applying the ethics of AI: a systematic review of tools for developing and assessing AI-based systems. Artif. Intell. Rev. 57:110 (2024).

examination of the ethical implications of AI, becomes clear that this is a fundamental question about the kind of future want to live in rather than just a technological one. Thus, the paper examines the global impacts and evolving trends in AI ethics, scruting the balance between technological advancement and moral responsibility, as well as the role of international cooperation in addressing these issues. Fig. 2 portrays the branches of AI which have a multitude of applications today.
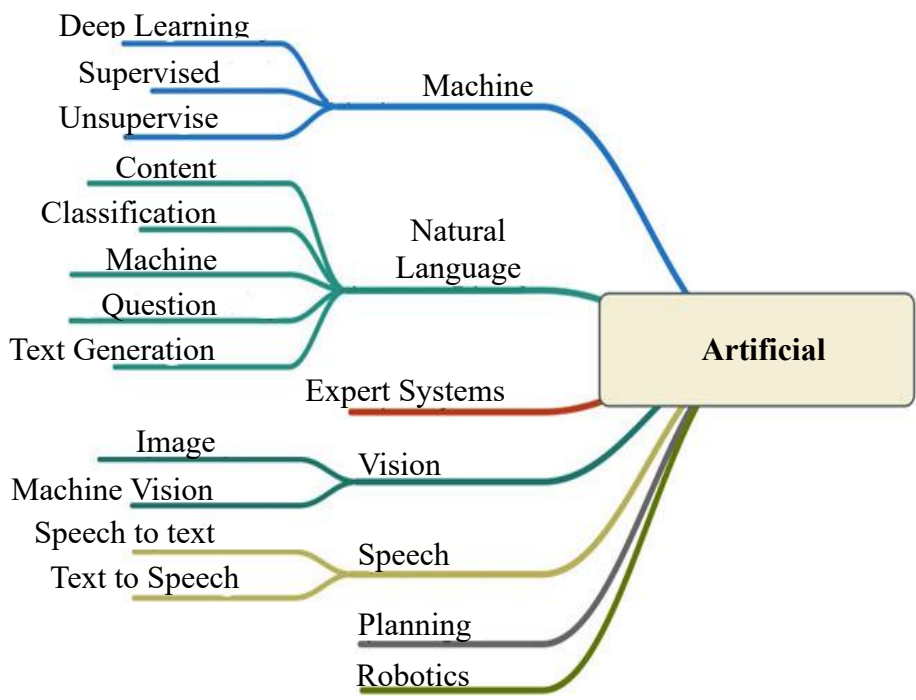
Fig 2: AI branches

In the twenty-first century, artificial intelligence has expanded significantly and is becoming more and more institutionalized where the application of AI-based technology in society at this critical juncture should support democratic principles like freedom, equality and transparency rather than work against them.[5] The focus of

---

[5] Aggarwal, C.C., Artificial Intelligence: A Textbook. Springer Nature Switzerland. ISBN 978-3-030-72356-9. (2021); Zeng, X. and Long, L. Introduction to Artificial Intelligence. In: Long, L. and Zeng, X. (Eds.). Beginning Deep Learning with Tensor Flow. Apress Berkeley, CA. ISBN 978-1-4842-7914-4. (2022).

artificial intelligence (AI) is increasingly on creating intelligent systems that can work well with humans which includes finding innovative approaches to create scalable, interactive platforms that allow humans to train robots. It will examine how these moral standards affect the application of AI technology and will draw attention to the changing discourse and trends in this area.

The Importance of Ethics in AI

Strong ethical frameworks are required to guarantee that AI technologies be created and used properly, given the technology's explosive expansion of AI.[6] The ethical concerns of artificial intelligence (AI) have drawn attention from a variety of communities as it becomes more and more prevalent in daily lives and advanced a multidimensional, multilevel conceptualization of trust in AI and examined the relationship between trust and ethics.[7] In addition, the ethical aspect of the creation and application of novel robotics and AI innovations, as well as their influence on contemporary society was well assessed.[8] The ethical considerations in AI address a wide range of issues, including:

- **Bias and Fairness**: AI systems can perpetuate or even exacerbate existing biases present in training data, leading to unfair outcomes. Ensuring fairness and eliminating bias is crucial for maintaining social justice. Bias in AI can occur at various stages of the AI development process, from data collection to algorithm design and deployment. Key sources of bias include data bias, algorithmic bias, human bias, and feedback loops. Data bias arises when the training data used to build AI models is skewed or unrepresentative, reflecting historical inequalities or flawed data collection methods.[9] The structure and design of algorithms can give rise to algorithmic bias, which, if left unchecked, may benefit particular groups. Developers and data scientists contribute human bias into the AI development process by bringing their own biases and preconceptions to the table. These biases affect decisions about algorithmic choices, data selection, and labelling. When AI systems engage with people, feedback

[6] Vyhmeister, E., Castane, G., Östberg, PO. And Thevenin, T., A responsible AI framework: pipeline contextualisation. AI Ethics. 3:175-197. (2023).

[7] Choung, H., David, P. and Ross, A., Trust and ethics in AI. AI & Soc. 38733-745. (2023).

[8] Hauer, T., Importance and limitations of AI ethics in contemporary society. Humanit. Soc. Sci. Commun. 9:272. (2022).

[9] Min, Alfonso., Artifical Intelligence and Bias: Challenges, Implications, and Remedies. Journal of Social Research. 2:3808-3817. (2023).

loops arise, producing biased outputs that intensify and reinforce biases in subsequent inputs. Therefore, it is imperative to address bias and fairness in AI to enable the appropriate development and application of AI technology.

- **Transparency and Explainability**: AI algorithms are opaque due to their complexity. Increasing openness and providing justification for AI judgements are crucial for fostering trust and accountability. Transparency and explainability have been emphasized in recent research as crucial AI system quality needs.[10] Gaining a thorough understanding of the mechanisms underlying the decision-making processes of autonomous systems, health care, finance and other critical domains, is becoming increasingly important as AI technologies become more sophisticated and seamlessly integrated which emphasizes how crucial it is to adhere to the concepts of explainability and transparency.[11]

- **Privacy and Security**: AI technology frequently uses large data sets are, which raises questions regarding data security and privacy. In order to stop abuse and security breaches, personal information must be protected at any cost. One of the most frequent worries people have regarding AI technologies is privacy.[12] The epistemic advantage that people typically have with regard to information about themselves and the kind of control over information about oneself that people rely on in regular interpersonal interactions is just two of the elements of privacy that AI systems do threaten.[13]

- **Autonomy and Control**: The increasing autonomy of AI systems poses questions about control and decision-making. Ensuring human oversight and control is necessary to avoid unintended consequences.

- **Accountability and Responsibility**: For millennia, autonomy has been seen as essential to both justice and well-being and it has been at the centre

---

[10] Balasubramaniam, N., Kauppinen, M., Hiekkanen, K. and Kujala, S., Transparency and Explainability of AI Systems: Ethical Guidelines in Practice. In: Gervasi, V., Vogelsang, A. (Eds.). Requirements Engineering: Foundation for Software Quality. REFSQ 2022. Lecture Notes in Computer Science. Vol. 13216. https://doi.org/ 10.1007/ 978-3-030-98464-9_1. (2022).

[11] Manure, A. and Bengani, S., S.S., Transparency and Explainability. In: Manure, Avinash., Bengani, Shaleen., Saravanan, S. (Eds.). Introduction to Responsible AI. Apress, Berkeley, CA. ISBN 978-1-4842-9981-4. (2023).

[12] Elliott, David and Soifer, Eldon., AI Technologies, Privacy, and Security. Frontiers in Artificial Intelligence. 5:826737. doi: 10.3389/frai.2022.826737. (2022).

[13] Oseni, Ayodeji., Moustafa, Nour., Janicke, Helge., Liu, Peng., Tari, Zahir., Rmit., Vasilakos, Athanasios., Security and Privacy for Artificial Intelligence: Opportunities and Challenges. *J. ACM*. 37(4): 111:35. https://arxiv.org/pdf/2102.04661. Accessed on 30.07.2024. 2020.

of moral and political thought. This view is supported by psychological research, which offers concrete proof that autonomy is essential for psychological wellness, personal development and motivation.[14] If responsible AI is to truly assist humanity, it will need to comprehend human autonomy (as opposed to just machine autonomy) and be able to build for it effectively.[15] In order to allocate accountability and handle liability concerns, precise protocols are required.

Global Ethical Guidelines for AI

To address ethical guidelines, nations and organizations have established ad hoc expert groups on AI, which are frequently tasked with creating policy texts. These committees include the UK House of Lords' Select Committee on Artificial Intelligence, the Advisory Council on the Ethical Use of Artificial Intelligence and Data in Singapore, the High-Level Expert Group on Artificial Intelligence appointed by the European Commission, and the expert group on AI in Society of the Organization for Economic Co-operation and Development (OECD).[16] These committees have reportedly produced or are in the process of preparing reports and advice materials on AI as part of their institutional assignments. 25 November 2021 Culture and Education. In a historic agreement, all of the member states of the UN Educational, Scientific, and Cultural Organization (UNESCO) established the shared values and tenets required to guarantee the sound growth of artificial intelligence (AI) by 2021. The Organisation for Economic Co-operation and Development (OECD) AI principles encourage the creative, reliable and human rights and democratic values-abiding use of AI.[17] OECD established AI criteria that are realistic and adaptable enough to last over time, and they were adopted in May 2019 which is updated in May 2024.

---

[14] Calvo, Rafael., Peters, Dorian., Vold, Karina. and Ryan, Richard., Supporting Human Autonomy in AI Systems: A Framework for Ethical Enquiry. In: Burr, C. and Floridi, L. (Eds.), Ethics of Digital Well-Being. Philosophical Studies Series 140, https://doi.org/10.1007/978-3-030-50585-1_2 10.1007/978-3-030-50585-1_2. (2020).

[15] Adam, Martin., Diebel, Christopher., Goutier, Marc. And Benlian, Alexander., Navigating autonomy and control in human-AI delegation: User responses to technology- versus user-invoked task allocation. Decision Support Systems. 180:114193. (2024).

[16] Jobin, Anna., Ienca, Marcello and Vayena, Effy., The global landscape of AI ethics guidelines. Nature Machine Intelligence. 1:389-399. (2019).

[17] OECD., Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Accessed on 30.07.2024. (2024).

Ethical Challenges and Debates

Artificial intelligence (AI) raises many complicated and nuanced ethical questions and controversies. The economic and employment implications of AI are crucial topics for ethical discourse.[18] AI-driven automation poses a danger to many employments, especially those that involve routine tasks. Although artificial intelligence (AI) has the potential to generate new employment opportunities, there is disagreement about how to handle the transition and guarantee that the advantages of AI are shared fairly throughout society. Other ethical conundrums are related to security and improper usage of AI technologies. AI can be used maliciously to create autonomous weapons, cyber-attacks and deep fakes, among other things.[19] Determining rules and measures to stop AI abuse while encouraging creativity is an important topic of discussion today. Deep ethical concerns are raised by the long-term effects of AI development, especially the possibility of superintelligent AI.[20] These concerns include the possibility of losing control over highly developed AI systems and the requirement for international collaboration to guarantee that AI is created and applied in ways that are advantageous to all people. These difficulties show that, in order to successfully negotiate the intricacies of AI research and deployment, a multidisciplinary strategy involving ethicists, technologists, policymakers, and the general public is required.

Even with the development of ethical principles, there are still a number of issues and disagreements. AI systems can reflect and amplify societal biases present in training data. Facial recognition technology, for instance, has come under fire for having greater error rates for particular demographic groups.[21] Diverse and representative data are needed to address prejudice and continual attempts to detect and reduce bias in AI models are also necessary. Many AI algorithms are "black box" systems, meaning that deciphering and justifying their

---

[18] Bankins, S. and Formosa, P., The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work. J. Bus. Ethics. 185:725-740. (2023).

[19] Yamin, Muhammad Mudassar., Ullah, Mohib., Ullah, Habib. and Katt, Basel., Weaponized AI for cyber-attacks. Journal of Information Security and Applications. 57:102722. (2021).

[20] Rayhan, Swajan., AI Superintelligence and Human Existence: A Comprehensive Analysis of Ethical, Societal, and Security Implications. 10.13140/RG.2.2.12922. 26561. Accessed on 30.07.2024. (2023).

[21] Birhane, Abeba., The unseen Black faces of AI algorithms. Nature. 610:451-452. (2022).

judgements is challenging.[22] This lack of openness has the potential to erode responsibility and confidence. The development of interpretable machine learning models and post-hoc explanation approaches are two strategies that researchers and practitioners are pursuing to improve the explainability of AI systems. The use of massive datasets by AI presents serious privacy and security issues and thus it is essential to guarantee data safety and stop unwanted access. To improve privacy in AI systems, strategies like federated learning and differential privacy are being investigated.[23] Concerns regarding human oversight and control are raised by the growing autonomy of AI systems, such as drones and self-driving cars. To avoid unforeseen effects and guarantee safety, it is crucial to establish explicit norms for human-AI interaction and decision-making.[24] It's difficult to assign blame for AI decisions. Assigning blame in situations of damage or malfunction can be difficult, as these problems must be addressed by legal and regulatory frameworks that clearly define responsibilities and lines of accountability.

Emerging Trends in AI Ethics

Emerging trends and new ethical issues accompany the advancement of AI technologies. The increased awareness of the need to handle the complex issues raised by AI technology is reflected in the emerging trends in AI ethics.[25] These patterns demonstrate the proactive measures that academics, businesses and governments are taking to guarantee that AI advances in ways that are advantageous and fair. Transparency in AI systems is becoming more and more in demand.[26] AI models that are comprehensible and explicable are being

---

[22] Blouin, Lou., AI's mysterious 'black box' problem, explained. The Regents of the University of Michigan. https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained. Accessed on 30.07.2024. (2023).

[23] Wei, K., Li, Jun., Ding, Ming., Ma, Chuan., Yang, Howard H., Farokhi, Farhad., Jin, Shi., Tony, Q.S., and Poor, H. Vincent., Federated Learning with Differential Privacy: Algorithms and Performance Analysis. IEEE Transactions on Information Forensics and Security. 15:3454-3469. (2020).

[24] de Almeida, P.G.R., dos Santos, C.D. and Farias, J.S., Artificial Intelligence Regulation: a framework for governance. Ethics Inf. Technol. 23:505-525. (2021); Cha, S., Towards an international regulatory framework for AI safety: lessons from the IAEA's nuclear safety regulations. Humanit. Soc. Sci. Commun. 11:506. (2024).

[25] Zhang, Yi., Wu, Mengjia., Tian George Yijun., Zhang, Guangquan., Lu, Jie., Ethics and privacy of artificial intelligence: Understandings from bibliometrics. Knowledge-Based Systems. 222:106994. (2021).

[26] Buijsman, S., Transparency for AI systems: a value-based approach. Ethics Inf. Technol. 26:34. https://doi.org/10.1007/s10676-024-09770-w. (2024).

promoted by stakeholders. The creation of fresh methods and resources that enable people to comprehend the decision-making processes of AI systems is one aspect of this trend.[27] Transparent AI makes it possible for decisions made by the system to be audited, contested and helped to establish confidence. Efforts address biases in AI are becoming more sophisticated as researchers are developing methods to detect and mitigate biases in training data and algorithms where this trend is crucial for creating AI systems that provide fair and equitable outcomes across different demographic groups.[28] There is also a growing emphasis on inclusive datasets that represent diverse populations.

There is a trend toward creating AI systems that can function without jeopardizing personal privacy as worries about data privacy grow. Methods like homomorphic encryption, federated learning and differential privacy are becoming more and more popular.[29] By using these techniques, AI is able to learn from data while safeguarding private data, meeting legal requirements and addressing privacy concerns. Governments and institutions are realizing that strong regulatory laws and governance structures are necessary for artificial intelligence. There is a movement to create rules, regulations and legislation to guarantee moral AI research and application.[30] These initiatives include establishing certification programs, setting up AI ethics boards, and enforcing adherence to moral guidelines. The use of AI for social benefit is becoming more and more popular and AI is being used in this trend to tackle urgent societal issues like disaster relief, healthcare, education and environmental sustainability. AI for social good highlights AI is beneficial effects on society and encourages the creation of applications that help marginalized and

---

[27] de Bruijn, Hans., Warnier, Martijn., Janssen, Marijn., The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly*. 39(2):101666. (2022).

[28] Shah, M. and Sureja, N., A Comprehensive Review of Bias in Deep Learning Models: Methods, Impacts, and Future Directions. Arch. Computat. Methods. Eng. https://doi.org/10.1007/s11831-024-10134-2. Accessed on 30.07.2024. (2024).

[29] Sébert, A.G., Checri, M., Stan, O., Sirdey, R. and Gouy-Pailler, C., Combining homomorphic encryption and differential privacy in federated learning. 20th Annual International Conference on Privacy, Security and Trust (PST), Copenhagen, Denmark, 2023, pp. 1-7, doi: 10.1109/PST58708.2023.10320195. (2023).

[30] Francés-Gómez, P., Ethical Principles and Governance for AI. In: Lara, F. and Deckers, J. (Eds.) Ethics of Artificial Intelligence. The International Library of Ethics, Law and Technology, vol 41. Springer, Cham. https://doi.org/10.1007/978-3-031-48135-2_10. (2023).

underprivileged groups.[31] AI development is progressively using human-centred design ideas. The needs, values and welfare of users and stakeholders are given priority in this trend. The goal of human-centred AI is to develop systems that respect user autonomy, advance human capabilities and offer fulfilling user experiences.[32]

A variety of disciplines, including computer science, philosophy, law, sociology and psychology must contribute to the discussion of AI ethics.[33] The trend in addressing the intricate ethical problems related to artificial intelligence is to collaborate across disciplines. Working together ensures the consideration of various viewpoints during the development and implementation of AI technology. It is becoming more and more crucial to establish explicit responsibility and liability frameworks for AI systems. Determining who is accountable for the decisions and acts of AI systems is a topic that is becoming more and more popular, especially in high-stakes industries like finance, healthcare, and autonomous driving.[34] The trend including AI ethics into professional training programs, public awareness initiatives and academic curriculum are commendable which encourages responsible AI practices among stakeholders who are aware of ethical considerations. These new developments show how seriously the ethical issues raised by AI are being taken. Stakeholders want to guarantee that AI technologies advance in a manner consistent with societal demands and human values and they do this by emphasizing openness, fairness, privacy, governance and collaboration.

AI Superintelligence and Human Existence

AI superintelligence holds significant implications for human existence as a possible kind of intelligence that transcends human cognitive capacities in all disciplines.[35] Such superintelligent systems

[31] Cowls, Josh., 'AI for Social Good': Whose Good and Who's Good? Introduction to the Special Issue on Artificial Intelligence for Social Good. Philosophy & Technology. 34:01-05. 10. (2021).

[32] Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C. and Xu, W., Six Human-Centered Artificial Intelligence Grand Challenges. International Journal of Human–Computer Interaction. 39(3):391-437. (2023).

[33] Kazim, Emre and Koshiyama, Adriano Soares., A high-level overview of AI ethics. Patterns. 2(9): 100314. (2021).

[34] Novelli, C., Taddeo, M. and Floridi, L., Accountability in artificial intelligence: what it is and how it works. AI & Soc. https://doi.org/10.1007/s00146-023-01635-y. Accessed on 30.07.2024. (2023).

[35] Mercer, C. and Trothen, T.J., Superintelligence: Bringing on the Singularity. In: Mercer, Calvin and Trothen, Tracy J. (Eds.) An Introduction to Biohacking, Artificial

could usher in revolutionary shifts, presenting enormous hazards as well as hitherto unheard-of opportunities. Potentially solving major global challenges with superintelligent AI is one of the most optimistic scenarios. An AI with this kind of computing capacity may improve resource management, speed up scientific research and find answers to urgent problems like poverty, illness and climate change. This might usher in a new era of sustainability, health and prosperity and greatly raise people's standard of living everywhere. But there are existential concerns associated with the development of superintelligent AI. The alignment problem - making sure that a superintelligent AI's objectives and behaviour are in line with human ideals and welfare - is one of the main causes for concern. Even though they make sense to misaligned AI, these goals could have disastrous or detrimental effects on humanity. Robust and failsafe methods are required in this scenario to govern and direct the operations of superintelligent AI. Superintelligent AI has the potential to do unheard-of amounts of harm if it is weaponized or employed maliciously by people, groups, or governments.[36] To lessen these risks, it is essential to make sure that such potent technologies are created and used properly, with sufficient protections and control.

The ramifications of AI superintelligence on society and ethics are equally significant. Superintelligent systems replacing human labour could cause severe economic disruption and the need for new strategies in the areas of social welfare, education, and employment.[37] In addition, ethical questions like privacy, autonomy and the nature of consciousness are raised, upending preconceived notions about philosophy and morality. Furthermore, the development of superintelligent AI has the potential to drastically change our conception of what it is to be human and the human experience.[38] The idea that AI could become more intelligent than humans begs concerns about the nature of intelligence, our place in the universe and the future of our humanity. It might cause humans' identity and purpose

---

Intelligence, and Transhumanism. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-62359-3_10. Accessed on 30.07.2024. (2021).

[36] Choi, Charles Q., Superintelligent AI May Be Impossible to Control; That's the Good News Postcard from the 23rd century: Not even possible to know if an AI is superintelligent, much less stop it. https://spectrum.ieee.org/super-artificial intelligence. Accessed on 30.07.2024. (2021).

[37] Shen, Y. and Zhang, X., The impact of artificial intelligence on employment: the role of virtual agglomeration. Humanit. Soc. Sci. Commun. 11:122. (2024).

[38] Gruetzemacher, Ross and Whittlestone, Jess., The transformative potential of artificial intelligence. Futures. 135:102884. (2022).

to be redefined in a world where machines are more intelligent than humans. It is essential that the creation of AI superintelligence be treated with caution, foresight and a strong ethical foundation given these possible results. To build comprehensive rules and regulatory frameworks, governments, researchers and industry leaders must work together in collaborative international endeavours. These ought to be designed to make sure that the development of AI superintelligence puts the welfare of humanity, morality and safety of people first. Thus, AI superintelligence holds the potential to either greatly enhance or severely threaten human existence. The path we take in its development will determine whether it becomes a boon or a bane for humanity. As we stand on the brink of this technological frontier, a careful, deliberate and ethically informed approach is crucial to harnessing the benefits of AI superintelligence while safeguarding against its risks.

Ethical AI by Design

The idea of "Ethical AI by Design" is to incorporate moral issues right from the start of the development process where this method is creating AI systems with security, privacy, transparency and justice in mind, instead of addressing ethical concerns after the fact.[39] Ethical AI by design refers to the process of incorporating moral issues from the beginning into the creation and application of artificial intelligence systems which makes sure AI technologies respect human rights, values and social standards. It includes a comprehensive perspective that covers every stage of the lifecycle of artificial intelligence systems, from design to implementation to decommissioning. The fairness concept, which tries to stop bias and discrimination in AI systems, is essential to this strategy. In order to address any biases that may emerge, this calls for the use of representative and diverse data sets in addition to ongoing monitoring and updating of AI systems.[40] Another important component is transparency, which promotes the idea that users and stakeholders should be able to comprehend and explain AI systems. Users are able to understand how decisions are made and, if

---

[39] Prem, E., From ethical AI frameworks to tools: a review of approaches. AI Ethics. 3:699-716. (2023).

[40] de Manuel, A., Delgado, J., Parra Jounou, I., Ausín, T., Casacuberta, D., Cruz, M., Guersenzvaig, A., Moyano, C., Rodríguez-Arias, D., Rueda, J., & Puyol, A., Ethical assessments and mitigation strategies for biases in AI-systems used during the COVID-19 pandemic. Big Data & Society. 10(1). https://doi.org/10.1177/2053951 7231179199. Accessed on 30.07.2024. (2023).

required, challenge or contest them because of this transparency, which promotes trust and accountability.

Developers are responsible for making sure that privacy laws and regulations are followed when collecting, processing and storing personal data.[41] While still enabling AI systems to operate efficiently, methods like data anonymization, encryption and differential privacy can help protect user information. Furthermore, ethical AI by design highlights the significance of responsibility, which implies that entities and people in charge of AI systems need to be recognized and held accountable for their deeds.[42] To direct the responsible development and application of AI, it is necessary to establish explicit governance frameworks and ethical standards. In order to create AI that is ethical, developers must consider the potential negative effects that their systems may have on employment, social interaction and general well-being. This progressive strategy promotes the development of AI technologies that improve human potential and benefit society. In order to ensure that AI systems are created with end users' wants and values in mind, ethical AI relies heavily on user-centric design. In order to get input and comprehend their viewpoints, this entails interacting with users and other stakeholders at various stages of the development process. AI systems may be made more approachable, intuitive and compliant with the moral standards of the communities they support by putting the human experience first.[43] In the end, ethical AI by design is about integrating moral values into AI systems from the ground up and this proactive approach encourages the development of AI systems that are reliable, equitable, and advantageous to society as a whole in addition to assisting in the prevention of harm and guaranteeing adherence to legal and regulatory requirements.

Conclusion

The field of AI ethics is vast and dynamic worldwide. People are hailing the idea and practice of creating computer systems, known as artificial intelligence (AI), as a continuous "revolution" that is

---

[41] Rahnama, Hossein and Pentland Alex Sandy., The New Rules of Data Privacy. Harward Business Review. https://hbr.org/2022/02/the-new-rules-of-data-privacy. Accessed on 30.07.2024. (2022).

[42] Stahl, B.C., Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems. Sci. Rep. 13:7586. (2023).

[43] Tsamados, A., Floridi, L. and Taddeo, M., Human control of AI systems: from supervision to teaming. AI Ethics. https://doi.org/10.1007/s43681-024-00489-4. Accessed on 30.07.2024. (2024).

transforming society and science in its entirety. As AI technologies progress, ethical issues need to be at the forefront of both development and application. It is imperative to tackle concerns like bias, transparency, privacy, autonomy and accountability to guarantee that AI systems maximize benefits while reducing drawbacks for society. Globally, nations and institutions are formulating moral protocols to address these issues. Even though there has been progress, more work is still required to properly develop and apply these rules. Prospective paths for furthering AI ethics include emerging concepts like inclusive participation, ethical AI by design and international collaboration. In the end, the ethical implications of AI raise deeper considerations about the kind of society wishes to create than merely technological ones. The use of artificial intelligence (AI) to build a more just, equitable and sustainable future by giving ethical issues first priority is the need of the hour. While there aren't many ethical concerns with the way autos and power plants are designed now thanks to AI, the direction AI algorithms are taking toward more human-like thinking signals inevitable problems. AI algorithms may take on social functions, requiring new design criteria like predictability and openness. It's possible that sufficiently general AI algorithms won't work in predictable scenarios, necessitating the development of new safety assurance mechanisms and artificial ethical considerations. AIs possessing enough sophisticated mental states or the appropriate kinds of states, will have moral standing and some may even be considered persons, albeit ones that may be substantially different from those that already exist and subject to other laws. Last but not least, the possibility of AI possessing superhuman intelligence and skills puts us in the unique position of having to develop an algorithm that produces super-ethical behaviour. Though these difficulties appear futuristic, it seems foreseeable that the world will face them and they do offer some ideas for current study directions.