

Ethical Artificial Intelligence Development: Meeting Obstacles and Shaping the Future

Executive Summary

The global Artificial Intelligence (AI) sector is poised for significant growth due to increasing investments, digital disruption, and the pursuit of competitive advantages. According to a market research firm, Grand View Research, Inc., it is projected that the market value for a certain industry would reach \$1,811.8 billion by the year 2030, with a compound annual growth rate of 37.3%. The unregulated proliferation of AI, however, presents ethical dilemmas including issues of bias, unemployment, breaches of privacy, and independent decision-making by AI systems, hence necessitating the development of ethical AI.

Consequently, this study aims to address the dearth of comprehensive research, established norms, and systematic protocols pertaining to the ethical use of AI. This issue is often seen to fall behind the advancements in technology and undermine the trust of the general population. The objectives of the effort include an evaluation of the current state of AI ethics, the identification of upcoming challenges, and an examination of the implications of ethical AI for both the corporate sector and society as a whole.

Contributing to ethical frameworks for AI, mitigating negative consequences, aligning AI systems with human values, fostering collaboration, and establishing a reliable AI system are all important endeavors. Hence, the examination of pertinent scholarly sources will clarify the expansion of AI, the ethical challenges it presents, and notable institutions and initiatives dedicated to fostering ethical practices in AI. Furthermore, it will highlight enduring concerns around bias, confidentiality, and responsibility.

During this project, the process of constructing conceptual definitions in response to significant issues (or challenges) involves the collaboration of six cycles of academics and industry professionals. These challenges include the absence of universally applicable standards, the implementation of ethical considerations in real-world scenarios, the advancement of AI governance, and the mitigation of prejudice. Focus groups have been shown to improve clarity and comprehension.

The anticipated outcomes encompass a thorough comprehension of AI ethics, the establishment of standardized practices, the acquisition of knowledge on mitigating bias, the development of a comprehensive ethical framework, the facilitation of interdisciplinary collaboration, the cultivation of heightened awareness, the generation of valuable policy insights, the enhancement of public trust, and the provision of contributions to the discourse on AI ethics.

Keywords: Artificial Intelligence (AI) Growth, Ethical AI Development, AI Ethics Challenges, AI Governance, Bias Mitigation in AI.

Background of the Project

The anticipated expansion of the worldwide Artificial Intelligence (AI) market in the forthcoming years may be attributed to augmented investments in AI technologies, the advent of digital disruption, and the pursuit of competitive advantage within the swiftly expanding global economy. Grand View Research, Inc (2023) has reported that the projected growth rate for the global AI market is estimated to be 37.3% on a compound yearly basis over the period from 2023 to 2030; the market is projected to have a value of \$1,811.8 billion by the year 2030.

AI has the capacity to enhance both corporate entities and societal structures. However, the unregulated expansion of AI technology gives rise to ethical considerations (More et al., 2022). The ethical concerns included in this context are the potential for bias in AI algorithms, the potential for employment displacement, the infringement upon privacy, and the possibility of autonomous systems making detrimental judgments. Due to the gravity of these challenges, it is imperative to prioritize the development and implementation of ethical and accountable AI (Raji et al., 2020).

The problem at hand

The crux of the matter is to the absence of thorough research, established norms, and standardized protocols that guarantee responsible and ethical implementations of AI. The advancement of technology often outpaces the development of ethical frameworks, leading to unanticipated consequences and a loss of public confidence (Munoko et al., 2020). Consequently, an endeavor is required that seeks to effectively address the issue by comprehensive research and providing robust recommendations for reducing the disparity.

The Project's Objectives

To facilitate the adoption and advancement of ethical AI, it is crucial to foster a collaborative environment among academic institutions, corporations, and governmental entities. Therefore, the primary objectives of this research endeavor are to determine the present state of AI ethics, to identify prospective developing difficulties, and to analyze the possible ramifications of ethical AI on both the corporate sector and the wider society landscape.

Project's Importance

This research project is significant because it will contribute significantly to the progress of ethical frameworks for AI. As a result, it will help to mitigate any negative repercussions while also facilitating the alignment of AI technology with human values and ethical norms. The fundamental of this research is to foster collaboration and allow the sharing of best practices, with the ultimate goal of developing a more responsible and dependable AI ecosystem.

Literature Review

AI has seen significant expansion and widespread use across several academic fields, leading to profound changes in both industrial sectors and the daily lives of individuals (Makridakis, 2017). Nevertheless, the fast progress of technology has introduced significant ethical problems that need thorough evaluation (More et al., 2022). Consequently, several notable examples of organizations focused on AI ethics include the AI Ethics Institute, the AI Ethics Lab, and the AI Now Institute. These organizations engage in research activities, provide valuable resources, and actively advocate for the ethical use of AI. Prominent technology firms have established research initiatives and teams dedicated to the study of AI ethics (Shneiderman, 2020). The AI Ethics team at Google and the Responsible AI team at Facebook are actively engaged in discourse and developing remedies to tackle ethical concerns. In order to tackle the ethical implications of AI, regulatory bodies such as the European Union (EU) and other national governments have implemented legislative measures. For instance, the proposal put up by the EU regarding the AI Act, together with similar legislative initiatives seen elsewhere, signifies the increasing global dedication towards the regulation of ethical considerations pertaining to AI (COM, 2021).

Currently, among the AI ethics, the issue of bias in AI systems has garnered significant attention (Mehrabi et al., 2021). AI systems, often trained on data that exhibits bias, have the potential to perpetuate and exacerbate societal biases. Scholars have been diligently engaged in developing methodologies to identify, alleviate, and preempt bias inside AI systems (Yu, 2020). However, the ongoing debates around the definition and evaluation of AI impartiality continue.

Moreover, the increasing reliance of AI systems on extensive datasets has led to the emergence of privacy concerns (Reddy et al., 2020). Striking a balance between data-driven innovation and the protection of individual privacy rights continues to pose a significant challenge. Considerable advancements are still required in the realm of formulating and implementing strategies and regulations aimed at safeguarding privacy in the context of AI (Nordström, 2022).

Moreover, the task of maintaining accountability in AI decision-making is a formidable problem (Enarsson et al., 2022). Academics and regulatory bodies are actively engaged in efforts to enhance the transparency and comprehensibility of AI systems. This encompasses the use of interpretability tools, the implementation of auditing methods, and the establishment of standards for the development of explainability in AI (Hamon et al., 2020). However, more investigation is necessary within this particular setting.

In summary, the present condition of AI ethics is marked by significant progress, obstacles, and an increasing recognition of the need for ethical AI research and implementation. The landscape of AI ethics is being influenced by ethical norms and conventions, alongside continuous research efforts and legislative initiatives. Nevertheless, the work of tackling concerns pertaining to discrimination, justice, privacy, and accountability remains a complex and ever-evolving endeavor that need continuous multidisciplinary cooperation between academic scholars and corporate stakeholders. To the best of our knowledge, there are four notable challenges that must be addressed in order to achieve success in this context.

- i. One of the primary challenges in the field of AI ethics is to the lack of universally applicable rules and regulations. Developing a cohesive framework for the ethical advancement of AI is challenging (Pedro, 2019), for instance, due to the divergent ethical norms and practices seen across different businesses and geographical locations.
- ii. Implementing ethical concepts in practical settings remains a formidable task in contemporary society. When making decisions, developers and organizations are faced with the need to carefully assess the balance between moral concerns and other practical factors (Anshari et al., 2022), for example, performance, budgetary limitations, and project deadlines.
- iii. The topic of AI governance is of significant academic interest and has garnered considerable attention in recent years. The development of effective governance for AI technology, including the design and enforcement of policies, is currently under progress. Legislative and oversight challenges arise due to the expeditious pace at which legislators endeavor to keep pace with the advancement of AI (Debbarma, 2023).
- iv. One way to address and mitigate bias is via the process of dispelling it. Despite extensive efforts, the mitigation of bias in AI systems continues to be an ongoing challenge. Developing algorithms that are fair and unbiased while considering diverse demographic and cultural characteristics is a significant challenge (Wang, 2020).

Methodology

The project will be finished in a total of six cycles. The first four rounds of the study will serve to elucidate the desired definitions and scope of the theoretical notions, as seen in Figure 1. During the first and subsequent rounds, the individuals involved will consist of scholars in the academic field and professionals with practical experience, respectively. During the selection process, careful consideration will be given to the inclusion of academicians and scholars from diverse countries who specialize in the field of AI Ethics and have already published relevant research in esteemed journals. Additionally, practitioners with a minimum of five years of experience as senior members of management in reputable AI companies will be chosen from various countries. In order to conduct a complete evaluation, it is recommended to obtain information from a minimum of 20 participants throughout the first and subsequent rounds (Shetty, 2019). Based on the data collected, this study aims to give conceptual definitions that might effectively address various difficulties.

Challenge 1: The absence of universally applicable laws and regulations.

RQ1: What are the primary determinants behind the absence of universally applicable laws and regulations in the realm of AI ethics, and how do divergent ethical standards and practices across various organizations and geographical areas impact the establishment of a cohesive ethical framework for AI?

Challenge 2: Utilizing Ethical Principles in Practical Scenarios.

RQ2: In the context of AI development, developers and organizations grapple with ethical considerations while making judgments. **What are the inherent trade-offs that must be navigated between ethical**

considerations and pragmatic issues such as performance, financial constraints, and project schedules?

Challenge 3: The evolution of AI governance and policy.

RQ3: The expeditious advancement of AI has had a profound impact on the existing legal and regulatory frameworks established to govern and supervise AI. **What are the primary obstacles and prospects associated with the development and implementation of robust AI governance, including the formulation and enforcement of policies?**

Challenge 4: The issue pertaining to the eradication of bias in AI systems.

RQ4: What are the current state-of-the-art methodologies and strategies used to mitigate bias in artificial AI systems? What measures can be taken to address concerns about the creation of unbiased and equitable algorithms that include diverse demographic and cultural characteristics?

Carpenter (2018) posits that mistakes in measuring might occur as a result of using intricate vocabulary or language, providing ambiguous replies, presenting biased arguments, and introducing modified notions. Consequently, a focus group with a minimum of five language specialists will be held during the third round. In the focus group meetings, participants will first be presented with the results of the initial two rounds. Furthermore, participants will be prompted with questions about the level of comprehension associated with the provided definition. What is your interpretation of this concept? "What are your thoughts upon encountering this specific inquiry? Furthermore, the participants' behavioral coding will be evaluated throughout the task, including noting instances of frowning or hesitation when being prompted to read the questions. In all, the ideas provided by the participants will be valuable as they effectively handle a range of issues, including ambiguity, uncertainty, leading questions, challenges, omissions, sensitivities, and question omissions. In the fourth round, the participants of the prior two rounds will get an email including the conceptual definitions that were generated in the preceding three rounds. The objective of this round will be to get further comments about the quality of definitions, as recommended by DeVellis (2012).

During the fifth and sixth rounds, the data collected in the preceding rounds will be used in order to accomplish the study objectives. Therefore, once again, individuals who are academically inclined and possess expertise in the field of AI will be recruited for their services. In due course, a forthcoming AI ethical framework is anticipated to undergo scrutiny by a collective body of project teams, practitioners, and professional advisors, operating within the realm of knowledge. The project's success will be contingent upon the favorable development of AI ethics. Furthermore, the project will be finalized by considering the general conversation.

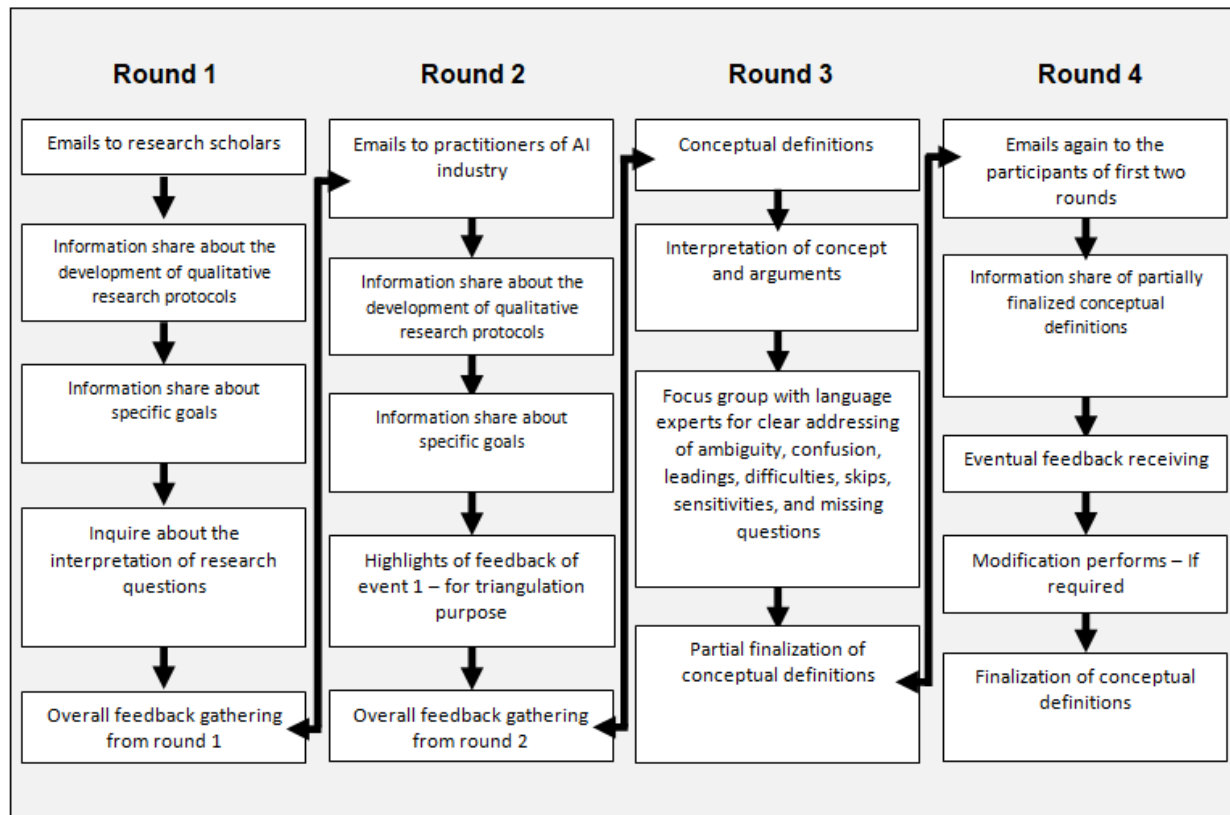


Fig.1. The first four rounds of the study serving to elucidate the desired definitions and scope of the theoretical notions

Expected Outcomes

This study endeavors to comprehensively comprehend the field of AI ethics by examining and analyzing prominent topics and concerns pertaining to the ethical use of AI. This necessitates a meticulous examination of prejudice, obstacles related to privacy, issues of responsibility, and governance complexities.

This research will enhance clarity and establish a uniform terminology throughout the field by providing conceptual definitions for essential ethical issues in the realm of AI. This will facilitate the development of a more comprehensive and universally accepted comprehension of the ethical considerations pertaining to AI.

The program is expected to provide advice and insights on the mitigation of bias in AI systems, including advanced methodologies and approaches. This will facilitate the development of algorithms that are more egalitarian and unbiased, since they consider a diverse range of demographic and cultural factors.

The project is anticipated to culminate in the development of a comprehensive framework for the ethical implementation of AI. The proposed framework will serve as a guiding structure for enterprises, governments, and academic institutions, facilitating the ethical and responsible use of AI technologies.

This study cultivates multidisciplinary collaboration by facilitating the interaction between business stakeholders and academic academics. This collaboration is intended to effectively tackle ethical concerns associated with AI.

Through the identification and exploration of ethical dilemmas associated with AI, as well as the provision of advice, it is anticipated that the program will foster a heightened recognition of the importance of ethical AI throughout diverse sectors, including academia, industry, and government.

This study provides valuable insights into the domain of AI governance and policies. It is expected to contribute to the understanding of the development and implementation of AI governance frameworks and policies. The findings of this research will assist policymakers in effectively responding to the rapid advancements in AI technology.

By fostering the establishment of ethical frameworks and standards, the aforementioned project has the capacity to enhance public trust in AI technology via the promotion of responsible and accountable AI implementations.

The anticipated outcome of the research is to enhance the existing corpus of knowledge in the field by making a valuable contribution to the scholarly discourse around the ethical considerations of AI.

This research has the potential to provide valuable insights to developers and corporations about the effective management of ethical considerations in the creation of AI, while simultaneously balancing performance objectives, budgetary limitations, and project timelines.

Further Discussion Required

- Research Plan and Timeline
- Budget and Resources
- Personnel (Collaboration and Partnerships)
- Equipment and Software
- Travel and Dissemination

References

Anshari, M., Syafrudin, M., Fitriyani, N. L., & Razzaq, A. (2022). Ethical Responsibility and Sustainability (ERS) Development in a Metaverse Business Model. *Sustainability*, 14(23), 15805.

COM, E. (2021). Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. Proposal for a regulation of the European parliament and of the council.

Debbarma, R. (2023). The Legal Framework And Challenges In Prosecuting Cybercrimes Including Hacking, Identity Theft, And Online Fraud. *resmilitaris*, 13(2), 7021-7048.

El Morr, C., Jammal, M., Ali-Hassan, H., & El-Hallak, W. (2022). Future Directions and Ethical Considerations. In *Machine Learning for Practical Decision Making: A Multidisciplinary Perspective with Applications from Healthcare, Engineering and Business Analytics* (pp. 449-460). Cham: Springer International Publishing.

Enarsson, T., Enqvist, L., & Naarttijärvi, M. (2022). Approaching the human in the loop—legal perspectives on hybrid human/algorithmic decision-making in three contexts. *Information & Communications Technology Law*, 31(1), 123-153.

Grand View Research. (2023). Artificial intelligence market worth \$1,811.75 billion by 2030. Artificial Intelligence Market Worth \$1,811.75 Billion By 2030. <https://www.grandviewresearch.com/press-release/global-artificial-intelligence-ai-market>

Hamon, R., Junklewitz, H., & Sanchez, I. (2020). Robustness and explainability of artificial intelligence. *Publications Office of the European Union*, 207.

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, 90, 46-60.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6), 1-35.

Munoko, I., Brown-Libur, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics*, 167, 209-234.

Nordström, M. (2022). AI under great uncertainty: implications and decision strategies for public policy. *AI & society*, 37(4), 1703-1714.

Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). Artificial intelligence in education: Challenges and opportunities for sustainable development.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020, January). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 33-44).

Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, 27(3), 491-497.

Shetty, S. (2018). Determining Sample Size for Qualitative Research: what Is the Magical Number?. Retrieved from. <https://interq-research.com/determining-sample-size-for-qualitative-research-what-is-the-magical-number/>.

Shneiderman, B. (2020). Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 10(4), 1-31.

Wang, R., Harper, F. M., & Zhu, H. (2020, April). Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Yu, P. K. (2020). The algorithmic divide and equality in the age of artificial intelligence. *Fla. L. Rev.*, 72, 331.