

## ETHICS AND PUBLIC TRUST IN AI GOVERNANCE: A LITERATURE REVIEW

Norainie Ahmad<sup>1\*</sup>, Muhammad Anshari<sup>2</sup>, Mahani Hamdan<sup>3</sup>, Emil Ali<sup>4</sup>

<sup>1</sup> Institute of Policy Studies, Universiti Brunei Darussalam  
Email: [norainie.ahmad@ubd.edu.bn](mailto:norainie.ahmad@ubd.edu.bn)

<sup>2</sup> Institute of Policy Studies, Universiti Brunei Darussalam  
Email: [anshari.ali@ubd.edu.bn](mailto:anshari.ali@ubd.edu.bn)

<sup>3</sup> Institute of Policy Studies, Universiti Brunei Darussalam  
Email: [mahani.hamdan@ubd.edu.bn](mailto:mahani.hamdan@ubd.edu.bn)

<sup>4</sup> Institute of Policy Studies, Universiti Brunei Darussalam  
Email: [emil.ali@ubd.edu.bn](mailto:emil.ali@ubd.edu.bn)

\* Corresponding Author

### Article Info:

#### Article history:

Received date: 25.06.2024

Revised date: 17.07.2024

Accepted date: 15.08.2024

Published date: 30.09.2024

#### To cite this document:

Ahmad, N., Anshari, M., Hamdan, M., & Ali, E. (2024). Ethics and Public Trust in AI Governance: A Literature Review. *International Journal of Law, Government and Communication*, 9 (37), 296-305.

DOI: 10.35631/IJLGC.937025

This work is licensed under [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)



### Abstract:

As AI systems become ubiquitous, policymakers and regulators must establish effective governance frameworks that fortify ethical values. Such a model should be value-driven, principle-based, and maintain ethical integrity and public trust. Challenges include generative AI's potential to exacerbate disinformation and other risks. This paper addresses integrating values into AI governance to foster ethical integrity and support innovation. Bibliometric analysis and case studies identify key value tensions and tradeoffs in AI regulation. We propose a governance framework that harmonizes values like transparency, accountability, and inclusivity with AI development. Collaborative governance and unlikely stakeholder coalitions can foster an innovation ecosystem prioritizing ethical practices. The findings guide policy and industry actors navigating AI regulation to align development with social and ethical values.

### Keywords:

Artificial Intelligence, Governance, AI Regulation, Ethics, Public Trust, Innovation, Bibliometric Analysis

## Introduction

Rapid advancements in artificial intelligence (AI) are impacting practically every facet of contemporary life (Sheikh, 2020). AI systems are becoming smarter and more widespread. Various areas use them, including customer service virtual assistants, self-driving cars, and

medical diagnosis in healthcare. AI has the potential to revolutionise organisations across various sectors, including public agencies (Elliott, 2019). AI is now part of decision-making processes, automating tasks that humans used to do (Anshari et al., 2023). For example, in healthcare, AI is assisting in diagnosing diseases; in finance, AI is automating stock trading; in transportation, AI is enabling autonomous vehicles; and in education, AI is personalising learning experiences. Furthermore, as AI capabilities increase, intelligent systems that enhance human cognitive abilities can have both positive and negative effects (Alam, 2021).

The exponential growth of AI capabilities has initiated a new era of technological revolution (Makridakis, 2017). AI systems have reached a point where they may outperform humans in various cognitive tasks, such as defeating world champions in chess and Go and analyzing vast datasets to discover complex patterns (Shamdi et al., 2022; Anshari et al., 2023a). While AI offers various benefits, such as enhanced efficiency, accuracy, and accessibility, it also presents significant ethical and practical challenges (Ochuba et al., 2024). Furthermore, AI poses substantial social and ethical dilemmas. Unregulated AI development has the potential to result in unforeseen outcomes, reinforce biases, compromise privacy, and potentially even pose significant risks to humanity (Khan, 2023). Artificial intelligence (AI) has the potential to significantly disrupt the workforce, affect privacy and security, and reinforce human biases (Anshari et al., 2021).

Effective governance frameworks are essential to align AI with human values and benefit society overall. This study examines the significance of AI governance and proposes possible governance methods. The use of AI has the potential to boost economic growth, decision-making, and technological advancements. On the other hand, AI that is misused or developed without adequate supervision could cause job loss, privacy violations, and even serious risks if intelligent systems aim to accomplish objectives that are not in line with reality. Because of these important issues, strong AI control has become a global issue that requires collaborative efforts from policymakers in setting regulations, ethicists in defining ethical boundaries, and technological experts in developing responsible AI systems. A thorough examination of the technological, ethical, legal, and societal aspects of AI research and deployment governance is necessary for a responsible navigation of this revolutionary age. This study aims to explore essential aspects of AI governance related to ethics and trust through an in-depth analysis of the literature and interviews with industry experts.

## Literature Review

An extensive literature analysis has examined various aspects of ethics and trust within the framework of AI governance. Researchers emphasise the necessity of establishing ethical rules for the development and implementation of AI systems. These criteria should include transparency, accountability, fairness, and human oversight (Akinrinola et al., 2024). This is important as AI systems grow increasingly sophisticated and widespread, in order to ensure that they align with human values and contribute to the overall welfare of society. Achieving a balance between ethical standards of transparency, accountability, fairness, and supervision by humans and technology progress brings challenges (Diakopoulos, 2020). Research indicates that too rigid regulations could prevent the ability to fully take advantage of the creative and adaptive capabilities of AI, consequently limiting innovation and advancement. This may prevent the extensive implementation and recognition of AI (Hagendorff, 2020). By fostering “trustworthy AI” through collaborative efforts including multiple stakeholders, ongoing evaluation, and adaptive governance, it is possible to effectively respond to the fast-paced

advancements in technology (AI, 2019). The task of balancing these divergent objectives remains to be an important priority in the discussion surrounding AI governance (Winfield et al., 2019).

### ***Ethics and AI***

Ethical considerations must be prioritised as AI systems become more autonomous and influential in high-stakes decision-making situations (Whittlestone et al. 2019). Establishing a governance framework based on explicit ethical standards is essential for ensuring that the development and implementation of AI technology give emphasis to human values such as fairness, accountability, and welfare. Additionally, governance policies should balance ethical restrictions with advancing AI capabilities (Dodda et al., 2021). Regulating AI too strictly could hinder the advancement of technology and prevent it from realising its immense potential benefits. Finding the right balance will necessitate careful consideration of both the risks and possibilities posed by AI.

With the growing complexity and popularity of AI systems, it is important to establish their development and implementation within robust ethical frameworks. Utilitarianism and deontology are two key ethical theories that offer valuable guidance for handling AI (Sommaggio & Marchiori, 2020). Utilitarianism, which was promoted by philosophers such as John Stuart Mill and Jeremy Bentham, highlights the concept that the morality of an action should be assessed based on its consequences and the degree to which it optimises the overall good of society, or utility (Zhang et al., 2022). From this perspective, the governance of AI should give priority to policies and practices that effectively utilise the technology's extensive ability to improve human well-being. This can be achieved through advancements in medical diagnostics, more effective allocation of resources, and the automation of hazardous or monotonous tasks (Taddeo & Floridi, 2018). Nevertheless, utilitarianism presents challenges as it may rationalise the prioritisation of the perceived larger good over the needs of specific individuals or minority groups.

On the other hand, Immanuel Kant's deontological ethics emphasises the fundamental rightness or wrongness of actions in light of general moral obligations and principles, such as respect for human dignity and autonomy as an individual. From this perspective, it is important for AI governance to establish resilient ethical measures in order to ensure fundamental human rights, even if this may result in sacrificing the goal of maximising overall social benefit. Deontological theorists argue that AI systems must never be allowed to make decisions that go against principles of justice and non-discrimination, irrespective of any potential advantages in efficiency (Jobin et al., 2019). This stance emphasises the ethical worth of each individual and the importance of not viewing people merely as means to an end. Rigid deontological standards, according to the opposition, can impede helpful developments and result in more difficult than desired outcomes.

Successfully adopting the complex ethical concerns of AI governance will likely require blending utilitarian and deontological viewpoints. Policymakers, ethicists, and technical experts must collaborate to develop governance frameworks that strike a balance between preserving morally unchanging values and utilising AI's transformational potential. It is necessary to provide explicit ethical principles, implement algorithmic auditing methods, and develop accountability mechanisms to ensure that AI systems preserve the rights of people, minimise harm, and advance the greater good. Moreover, building public trust requires

transparency and involving all relevant parties in decision-making. The general acceptance and use of AI depend on how people perceive its ethical implementation. By integrating key ethical theories, AI administrators can strive to reap the benefits of the technology while safeguarding essential human values.

### ***Public Trust and AI***

This literature study aims to analyse and combine many viewpoints from ethics, governance, and technological fields to evaluate potential AI governance frameworks in terms of their ability to maintain ethical standards and establish public confidence. These insights can be used together to establish solid governance methods that ensure AI benefits society as a whole while minimising challenges. Public trust is a fundamental component and objective of good AI governance. If many people perceive AI as a potential danger to humanity, it might significantly prevent the acceptance and implementation of innovative technologies based on AI. To foster trust, it is essential for governance mechanisms to give priority to ensuring the transparency, clarity, and accountability of AI systems.

Recent research emphasises the vital role of public trust in managing and implementing AI systems. The extent to which the public is willing to accept and use AI technologies will play an important part in determining their overall impact on society, especially as they become more integrated into critical decision-making processes and essential applications. Studies suggest that widespread adoption of advanced AI applications could face obstacles due to doubts about AI systems' transparency, accountability, and potential misuse (Hagendorff, 2020). On the other hand, it is needed to develop robust public trust in the safety, dependability, and compatibility of AI with human values in order to take full advantage of its potential.

Several academic articles highlight many key factors that play a role in establishing public confidence in AI. According to the High-Level Expert Group on AI, demonstrating transparency, clarity, and interpretability is the primary requirement for AI systems in 2019. Lack of transparency in the internal processes and decision-making procedures of AI systems affects the public's perception of their credibility and trustworthiness (AI, 2019). Researchers argued for the creation of explainable AI (XAI) methods that can offer people understandable explanations of how AI systems reached their results (Arrieta et al., 2020; Dwivedi et al., 2023). Moreover, having mechanisms for human oversight and the ability to challenge AI decisions are crucial for establishing trust and accountability (Jobin et al., 2019). Incorporating these concepts into the governance and design of AI systems could reduce the public's concerns over the potential for bias, error, and misuse.

Moreover, the research highlights the importance of continuous, collaborative efforts across various groups in order to foster public trust in AI. It is vital for policymakers, industry leaders, ethicists, and tech experts to collaborate in creating flexible governance structures that can adapt to rapid advancements in AI (Taddeo & Floridi, 2018). Adopting open communication, using inclusive decision-making processes, and demonstrating a commitment to integrating AI with society's values can all help to increase public confidence. Researchers believe that building trust is not a singular effort but rather a recurring process involving ongoing assessment, adaptation, and the involvement of the public. Making trust a core principle in AI governance enables policymakers to secure public acceptance of innovations that benefit society.

## Methodology

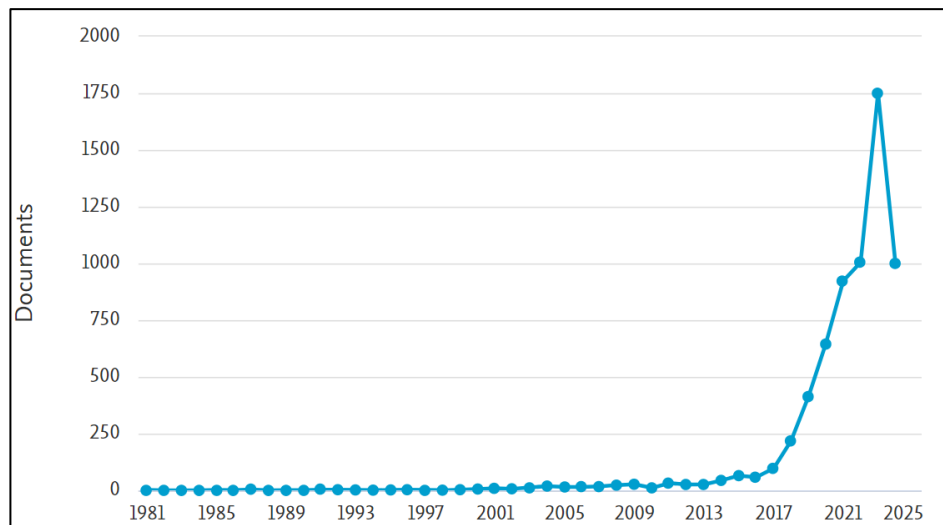
This study used a variety of research methods, mainly based on a thorough review of existing literature and an analysis of publication patterns. The literature review combined information from academic journals, policy papers, and industry reports to comprehensively explore the key issues and debates related to artificial intelligence governance. The literature covered various fields like computer science, ethics, law, and public policy to reflect the interdisciplinary aspect of the topic. Furthermore, a bibliometric analysis was performed with the Scopus database to pinpoint and evaluate significant published research related to AI governance. This numerical evaluation of the academic field offered practical insights into the main topics, partnerships, and areas needing more research in the current body of work. By combining the qualitative analysis of the literature review with the bibliometric mapping of research publications, this study aimed to give a full, evidence-based picture of how to govern AI technologies in terms of ethics, innovation, and trust. The results from this combined research approach guide the formulation of actionable suggestions for policymakers, industry leaders, and other important decision-makers involved in shaping the direction of AI.

## Analysis and Discussion

As artificial intelligence continues to advance at a rapid pace, the question of how these transformative technologies should be adopted and governed has become a critical concern for policymakers, industry leaders, and the general public. The study findings emphasise the need for a comprehensive and balanced approach to managing artificial intelligence, taking into account ethical considerations and factors related to trust. From the results of a bibliometric analysis using the keywords "artificial intelligence" AND "ethics" on all publications indexed in the Scopus database, there are a total of 6,499 publications under consideration. Interestingly, research on this topic began to be discussed in 1981 and has seen a significant increase from 2017 to the present (See Figure 1).

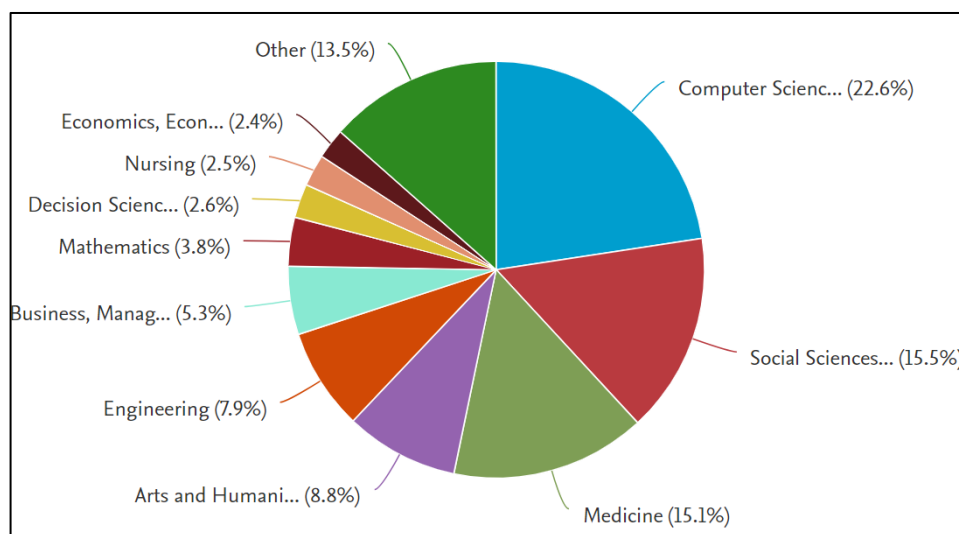
The subject areas that predominantly cover this topic are computer science at 23%, followed by social science at 15.5%. Other fields include medicine, arts & humanities, engineering, business & management, mathematics, and decision science (See Figure 2). This analysis highlights the growing interdisciplinary interest in the intersection of artificial intelligence and ethics, reflecting the expanding implications of AI technology in various fields. The substantial increase in publications since 2017 indicates a heightened awareness and scholarly attention towards the ethical considerations of AI, driven by advancements in AI technologies and their pervasive impact on society. The distribution of research across multiple disciplines underscores the multifaceted nature of ethical issues in AI, necessitating diverse perspectives and approaches to address them comprehensively.





**Figure 1: Documents by Year, for the Bibliometric Search on "Artificial Intelligence" AND "Ethics"**

Source: Authors' Findings



**Figure 2: Documents by Subject Area, for the Bibliometric Search on "Artificial Intelligence" AND "Ethics"**

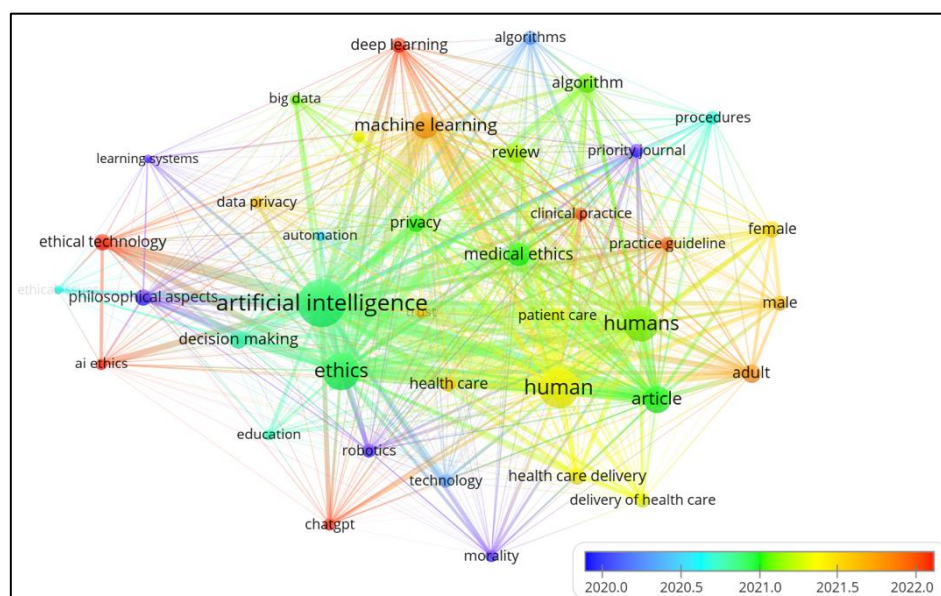
Source: Authors' Findings

Meanwhile, Figure 3 below illustrates the trends and research patterns on the topic of "artificial intelligence" AND "ethics" from the year 2020 onwards. In 2020, research topics visualized in blue extensively discussed themes such as morality, robotics, philosophical aspects, algorithms, and learning systems. Towards the end of 2020, the focus began to shift towards themes like ethical issues, procedures, automation, decision-making, and educational aspects.

In 2021, the intersection of AI and ethics predominantly covered themes like AI and privacy, medical ethics, healthcare and AI, healthcare delivery, and Big Data. Then, in 2022, the

research focus expanded to include discussions on ethical technology, machine learning, deep learning, data privacy, ChatGPT, and AI ethics.

This evolution in research themes highlights the dynamic nature of AI and ethics discourse, adapting to technological advancements and societal needs, in line with innovations in the field of AI and its related technologies. The early focus on foundational concepts such as morality and philosophical aspects set the stage for more applied concerns like ethical issues in automation and decision-making. As AI technologies became more integrated into critical areas like healthcare and privacy, research in 2021 reflected these pressing concerns. By 2022, the discourse had further evolved to tackle emerging technologies and their ethical implications, including sophisticated AI models like machine learning, deep learning, and AI-powered tools such as ChatGPT, emphasizing the ongoing need to address data privacy and ethical considerations in these advanced systems.



**Figure 3: Overlay Visualisation, for the Bibliometric Search on "Artificial Intelligence" AND "Ethics"**

Source: Authors' Findings

### ***Ethical Compliance***

The literature emphasises the need to base AI development and deployment on strong ethical frameworks. Ethical theories like utilitarianism and deontology provide important principles that should be included in AI governance strategies. A utilitarian approach would emphasise maximising the overall societal benefits of AI by harnessing its potential to enhance human welfare, while a deontological perspective would prioritise inviolable moral duties and individual rights, even if doing so comes at the expense of aggregate utility gains. Combining these ethical perspectives helps AI stay in line with core human values and advance the common good. However, the literature also discusses that overly restrictive ethical constraints could inadvertently stifle innovation and prevent societies from realising the full transformative potential of AI. Governance policies need to balance ethical boundaries with fostering the agility and creativity essential for ongoing progress. Building public trust through transparent, accountable, and inclusive decision-making processes is vital for navigating this delicate

balance. By embracing a holistic, ethically grounded approach to AI adoption, policymakers can harness the technology's immense benefits while mitigating its risks and earning the confidence of the broader populace.

### ***Public Trust***

Successful integration of AI depends not just on its technical capabilities but also on gaining the trust and acceptance of the general public. A key insight that emerges from the literature is the critical role of public trust in shaping the trajectory of AI adoption. Studies have shown that a lack of trust in the safety, reliability, and ethical alignment of AI systems could greatly limit their widespread use. This can happen even when the technologies provide significant functional advantages. Therefore, it is crucial to prioritise building strong public trust as a key foundation of any AI governance framework. This requires a comprehensive strategy that focuses on transparency, clarity, and accountability throughout the AI development process. In particular, it is essential to have mechanisms for human oversight, the capacity to question AI-based choices, and clear descriptions of how these systems reach their results. Without these protections, the public may perceive AI as a mysterious system—evident, untrustworthy, and possibly biased. By engaging in proactive communication and involving stakeholders, you can strengthen public trust by showing a sincere dedication to aligning AI with societal values and addressing valid worries. By prioritising trust as a fundamental factor, policymakers can facilitate the adoption of AI technologies that are widely accepted and used for the greater good.

### ***Potentials & Challenges***

The literature discusses the potential benefits and challenges of integrating artificial intelligence (AI) from ethical and trust-related perspectives. AI can greatly improve human well-being by automating risky, boring, or error-prone tasks, optimising resource distribution, and speeding up scientific and medical advancements. Moreover, well-designed AI systems can make fairer decisions than humans, reducing the chance of discrimination or unfair results. Moreover, improvements in explainable AI (XAI) can give users more insight into how these systems work and make decisions, building trust and allowing for better supervision.

Nevertheless, there are significant ethical and trust-related challenges associated with the integration of AI. If AI systems are trained on biased data or flawed algorithms, they may worsen historical inequalities and discrimination, going against fairness and equality principles. AI's heavy reliance on data also raises important issues regarding personal privacy, data ownership, and the potential misuse of personal information. As AI gains autonomy, determining moral accountability for its actions becomes harder, potentially damaging public trust. Furthermore, the idea of AI systems surpassing human abilities and acting against human values has sparked fears of severe threats to humanity. Lastly, the replacement of human jobs by AI-driven automation could worsen wealth gaps and social divisions without fair distribution of the benefits. Navigating these complex trade-offs will be crucial to ensuring AI is deployed in a manner that upholds ethical principles and earns the trust of the broader populace.

### ***Conclusion***

The governance of artificial intelligence (AI) is a challenging task that requires a thoughtful and detailed approach. Balancing innovation and ethical principles while building public trust and ensuring fair access to AI's benefits is essential for unlocking its transformative potential.



This study highlights the need for a collaborative approach involving multiple stakeholders in AI governance, incorporating ethical guidelines, enhancing transparency, and prioritising the creation of reliable AI systems that reflect human values. By considering these essential factors, policymakers, industry leaders, and researchers can collaborate to harness AI's potential for positive impact, leading to a fairer, more inclusive, and prosperous future for everyone.

### Acknowledgment

The authors would like to acknowledge the organisers of the 16th International Conference on Humanities and Social Sciences (16th ICHiSS 2024), National Defence University of Malaysia, for accepting this paper for presentation during the event.

### References

- Akinrinola, O., Okoye, C. C., Ofodile, O. C., & Ugochukwu, C. E. (2024). Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. *GSC Advanced Research and Reviews*, 18(3), 050-058.
- Alam, A. (2021, November). Possibilities and apprehensions in the landscape of artificial intelligence in education. In *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)* (pp. 1-8). IEEE.
- Anshari, M., Almunawar, M. N., Masri, M., & Hrdy, M. (2021). Financial technology with AI-enabled and ethical challenges. *Society*, 58(3), 189-195.
- Anshari, M., Hamdan, M., Ahmad, N., Ali, E., & Haidi, H. (2023). COVID-19, artificial intelligence, ethical challenges and policy implications. *Ai & Society*, 38(2), 707-720.
- Anshari, M., Almunawar, M. N., & Masri, M. (2023a). Modelling Autonomous Vehicle Safety in Road Scenarios Considering User Behaviour. In *The International Conference on Artificial Intelligence and Applied Mathematics in Engineering* (pp. 269-278). Cham: Springer Nature Switzerland.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, 58, 82-115.
- AI, H. (2019). High-level expert group on artificial intelligence. *Ethics guidelines for trustworthy AI*, 6.
- Dodda, S. B., Maruthi, S., Yellu, R. R., Thuniki, P., & Reddy, S. R. B. (2021). Ethical Deliberations in the Nexus of Artificial Intelligence and Moral Philosophy. *Journal of Artificial Intelligence Research and Applications*, 1(1), 31-43.
- Diakopoulos, N. (2020). Accountability, transparency, and algorithms. *The Oxford handbook of ethics of AI*, 17(4), 197.
- 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.
- Elliott, A. (2019). *The culture of AI: Everyday life and the digital revolution*. Routledge.
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and machines*, 30(1), 99-120.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, 1(9), 389-399.
- Khan, A. (2023). Harnessing the Power of AI: A Review of Advancements in Healthcare. *BULLET: Jurnal Multidisiplin Ilmu*, 2(3), 546-556.
- Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, 90, 46-60.

- Ochuba, N. A., Adewunmi, A., & Olutimehin, D. O. (2024). The role of AI in financial market development: enhancing efficiency and accessibility in emerging economies. *Finance & Accounting Research Journal*, 6(3), 421-436.
- Shamdi, W., Lai, D., Aziz, A. A., & Anshari, M. (2022). Artificial intelligence development in Islamic System of Governance: a literature review. *Contemporary Islam*, 16(2), 321-334.
- Sheikh, S. (Ed.). (2020). *Understanding the role of artificial intelligence and its future social impact*. IGI Global.
- Sommaggio, P., & Marchiori, S. (2020). Moral dilemmas in the AI era: A new approach. *Journal of Ethics and Legal Technologies*, 2(JELT-Volume 2 Issue 1), 89-102.
- Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, 361(6404), 751-752.
- Winfield, A. F., Michael, K., Pitt, J., & Evers, V. (2019). Machine ethics: The design and governance of ethical AI and autonomous systems [scanning the issue]. *Proceedings of the IEEE*, 107(3), 509-517.
- Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019, January). The role and limits of principles in AI ethics: Towards a focus on tensions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 195-200).
- Zhang, Z., Chen, Z., & Xu, L. (2022). Artificial intelligence and moral dilemmas: Perception of ethical decision-making in AI. *Journal of Experimental Social Psychology*, 101, 104327.