

# The Ethics of Artificial Intelligence: Challenges, Dimensions, And Prospects for Responsible Use

Ibrahim Nehad Ineizeh<sup>1\*</sup>

<sup>1</sup> School of Electronics, Qween's Belfast University. Belfast, UK.

## CHRONICLE

Article history:

Received: June, 25, 2025.

Received in revised format: August, 04, 2025.

Accepted: December, 06, 2025.

Available online: December, 31, 2025.

## Keywords:

*AI ethics, AI governance, fairness, accountability, privacy, GDPR, responsible AI.*

## ABSTRACT

The penetration of artificial intelligence in sensitive sectors such as health, finance, and public administration is advancing apace, giving rise to important ethical and governance problems. Due to fairness, accountability, transparency, and privacy issues (FAT/ML = Fairness Accountability-Transparency/Machine Learning), they are central in the responsible deployment of AI. In reaction, rules and laws like the European Union General Data Protection Regulation (GDPR) have been implemented to safeguard people's rights and to provide guidelines for the ethical use of AI. But the enactment of such laws also presents new challenges related to the availability of data, the accuracy of models, and the performance of systems. In this paper, we conduct a thorough characterization of AI ethics and governance, specifically in terms of fairness, accountability, and privacy, and discuss the effects, such as GDPR, on the performance and robustness of AI models. The results call attention to the costs and benefits of ethical adherence vs. technical performance, and to novel governance paradigms that strive to reconcile innovation and responsible AI use.

## الملخص

يتزايد انتشار الذكاء الاصطناعي في قطاعات حساسة كالصحة والمالية والإدارة العامة بوتيرة متسارعة، مما يُثير مشكلات أخلاقية وحكومية هامة. ونظرًا لقضايا العدالة والمساءلة والشفافية والخصوصية (العدالة والمساءلة والشفافية/التعلم الآلي)، تُعدّ هذه القضايا أساسية في الاستخدام المسؤول للذكاء الاصطناعي. واستجابةً لذلك، تم تطبيق قواعد وقوانين مثل اللائحة العامة لحماية البيانات (GDPR) للاتحاد الأوروبي لحماية حقوق الأفراد وتوفير إرشادات للاستخدام الأخلاقي للذكاء الاصطناعي. إلا أن سنّ هذه القوانين يطرح تحديات جديدة تتعلق بتوافر البيانات ودقة النماذج وأداء الأنظمة. في هذه الورقة، نقدم توصيفًا شاملاً لأخلاقيات الذكاء الاصطناعي وحكومته، وتحديدًا فيما يتعلق بالعدالة والمساءلة والخصوصية، وناقش تأثيرات قوانين مثل اللائحة العامة لحماية البيانات (GDPR) على أداء نماذج الذكاء الاصطناعي ومثابنتها. وتلفت النتائج الانتباه إلى تكاليف وفوائد الالتزام الأخلاقي مقابل الأداء التقني، وإلى نماذج الحوكمة الجديدة التي تسعى جاهدة للتوفيق بين الابتكار والاستخدام المسؤول للذكاء الاصطناعي.

## الكلمات الدالة:

*أخلاقيات الذكاء الاصطناعي، وحكومة الذكاء الاصطناعي، والإنصاف، والمساءلة، والخصوصية، واللائحة العامة لحماية البيانات، والذكاء الاصطناعي المسؤول.*

**JEL Classification:** O33, D83, K24, D63, L86, O38.

\* Corresponding author.

E-mail address: [iineizeh01@qub.ac.uk](mailto:iineizeh01@qub.ac.uk)  
<https://doi.org/10.70568/UJFI AI.2.2.12.25.2>

All rights reserved.

## 1. Introduction

Artificial Intelligence (AI) has rapidly evolved into one of the most transformative technologies of the twenty-first century, fundamentally reshaping economic systems, institutional practices, and everyday human interactions. From predictive analytics in finance to diagnostic support in healthcare and automated decision-making in public administration, AI systems are increasingly embedded in high-stakes environments where their outputs directly influence human welfare and societal outcomes (Duan et al., 2019; Green & Chen, 2021). This rapid diffusion has been driven by advances in data availability, computational power, and machine learning techniques, positioning AI not only as a tool for efficiency but also as a central driver of innovation and competitiveness across sectors (Taeihagh, 2021). However, this technological acceleration has outpaced the development of adequate ethical and regulatory frameworks, raising critical concerns about the societal implications of AI deployment (Renda, 2019).

A growing body of research highlights that AI systems are not neutral artifacts; rather, they reflect the assumptions, biases, and limitations embedded in their design, data, and implementation contexts. Issues such as algorithmic bias, opacity in decision-making (often referred to as the “black box” problem), lack of accountability, and risks to individual privacy have emerged as central challenges (Floridi et al., 2018; Mittelstadt et al., 2016). These concerns are particularly acute in sensitive domains where AI-driven decisions can reinforce existing inequalities or produce unjust outcomes (Varona & Suárez, 2022). While proponents argue that AI enhances objectivity and efficiency, critics emphasize that without careful oversight, such systems may institutionalize discrimination, undermine trust, and erode fundamental rights (Fletcher et al., 2021). This tension underscores the need for a critical examination of the ethical foundations underpinning AI technologies. In response to these challenges, governments and international organizations have begun to develop regulatory frameworks aimed at ensuring the responsible use of AI. Among these, the European Union’s General Data Protection Regulation (GDPR) represents a landmark effort to safeguard individual rights in the digital age, emphasizing principles such as data minimization, transparency, and accountability (Mitrou, 2018; Sartor & Lagioia, 2020). Although such regulations provide essential safeguards, they also introduce new complexities for AI development, including constraints on data accessibility, increased compliance costs, and potential trade-offs between ethical requirements and model performance (Alic, 2021). This raises an important question: to what extent can regulatory interventions balance the dual objectives of fostering innovation and ensuring ethical integrity?

The motivation of this study stems from the recognition that current debates on AI ethics often remain fragmented, addressing fairness, accountability, transparency, and privacy in isolation rather than as interconnected dimensions of a broader governance challenge. Moreover, there is limited critical analysis of how ethical and regulatory constraints concretely affect the technical performance and robustness of AI systems (Dwivedi et al., 2023; Mikalef et al., 2022). This paper seeks to bridge this gap by offering an integrated perspective that not only conceptualizes the core ethical principles of AI but also critically evaluates their practical implications in real-world applications. Ultimately, this research is driven by the need to move beyond normative discussions toward a more balanced and critical discourse that acknowledges both the promises and limitations of ethical AI frameworks. It argues that achieving responsible AI is not merely a technical or regulatory challenge but a socio-technical endeavor requiring collaboration among developers, policymakers, and society at large (Buhmann & Fieseler, 2021; Herrmann, 2023). Through this lens, the paper aspires to provide insights into how emerging governance paradigms can reconcile innovation with ethical responsibility, ensuring that AI technologies serve as instruments of societal benefit rather than sources of unintended harm.

## 2. Literature Review

In recent years, the ethics of AI has emerged as a central concern in both academic and policy-oriented debates, reflecting the growing recognition that technological advancement must be aligned with societal values and human rights. Early contributions to the field primarily focused on conceptualizing the ethical risks associated with algorithmic systems, including bias, opacity, and the erosion of accountability (Mittelstadt et al., 2016; Floridi et al., 2018). These studies established that AI systems, particularly those

based on machine learning, are inherently shaped by the data on which they are trained and the assumptions embedded within their design, thereby challenging the notion of technological neutrality. As a result, ethical considerations have increasingly been framed not as external constraints but as integral components of AI system development. A significant strand of the literature has examined the issue of fairness and bias in AI systems, emphasizing how algorithmic decision-making can reproduce or even amplify existing social inequalities. Empirical and theoretical studies demonstrate that biased datasets, flawed model design, and unrepresentative sampling can lead to discriminatory outcomes across various domains, including healthcare, employment, and criminal justice (Varona & Suárez, 2022; Pagano et al., 2023). While numerous technical solutions have been proposed (such as fairness-aware algorithms and bias mitigation techniques) scholars argue that these approaches often involve trade-offs between fairness and model accuracy, raising questions about how to balance ethical priorities with performance objectives (Le Quy et al., 2022). This tension highlights the complexity of operationalizing ethical principles within real-world AI systems.

Another key area of research focuses on accountability and transparency, particularly in the context of complex and opaque AI models. The so-called “black box” nature of advanced machine learning systems, especially deep learning, has raised concerns about the ability of stakeholders to understand, audit, and challenge automated decisions (Busuioc, 2021; Dwivedi et al., 2023). In response, the field of explainable artificial intelligence (XAI) has gained prominence, aiming to enhance the interpretability of AI systems without significantly compromising their performance (Das & Rad, 2020). However, existing studies suggest that explainability techniques may introduce additional layers of complexity and may not fully resolve issues of accountability, particularly in high-stakes applications where transparency is critical for legal and ethical oversight (Chinnaraju, 2025).

Privacy and data protection constitute another major dimension of AI ethics, given the reliance of AI systems on large-scale data collection and processing. The literature emphasizes that the use of personal and sensitive data raises significant concerns regarding consent, data security, and potential misuse (Duan et al., 2019; Radanliev et al., 2024). Regulatory frameworks such as the GDPR have been widely studied as mechanisms to address these concerns by enforcing principles such as data minimization, purpose limitation, and user consent (Mitrou, 2018; Sartor & Lagioia, 2020). Nevertheless, scholars note that compliance with such regulations can impose constraints on data availability and model training, potentially affecting the performance and scalability of AI systems (Alic, 2021). This has led to an ongoing debate about whether strict data protection laws may inadvertently hinder innovation. Beyond these core ethical dimensions, recent literature has expanded to address broader societal and governance implications of AI. Studies have explored issues such as the digital divide, labor market disruption, and the ethical risks associated with autonomous systems, including military applications (Helbing, 2018; Amoroso & Tamburrini, 2020). Furthermore, international organizations such as the OECD and UNESCO have advocated for the development of global ethical frameworks that promote human-centered AI and sustainable development. These initiatives reflect a growing consensus that effective AI governance requires not only technical solutions but also institutional coordination, policy integration, and cross-sector collaboration.

Despite the richness of existing research, a critical gap remains in integrating these diverse perspectives into a cohesive analytical framework. Much of the literature treats fairness, accountability, transparency, and privacy as distinct challenges, with limited attention to their interdependencies and combined impact on AI system performance. Additionally, there is insufficient empirical and conceptual exploration of the trade-offs between ethical compliance and technical efficiency. Addressing this gap is essential for advancing a more holistic understanding of responsible AI. Accordingly, this study builds upon the existing literature by offering an integrated and critical analysis of AI ethics and governance, with particular emphasis on the interplay between ethical principles, regulatory frameworks, and system performance.

### ***2.1 The Concept of Artificial Intelligence Ethics***

The concept of AI ethics has emerged as a critical field of inquiry aimed at guiding the design, development, and deployment of AI systems in ways that align with human values, fundamental rights, and societal well-being. At its core, AI ethics refers to a set of normative principles and frameworks that seek to ensure that intelligent systems operate in a manner that is fair, transparent, accountable, and respectful of human dignity (Floridi et al., 2018; Leslie, 2019). Unlike traditional technological ethics, which often focus

on the responsible use of tools, AI ethics addresses the unique challenges posed by systems capable of autonomous learning, adaptation, and decision-making in complex and often opaque environments. A defining feature of AI ethics is its interdisciplinary nature, drawing from philosophy, law, computer science, and social sciences to address the multifaceted implications of AI technologies. From a philosophical perspective, AI ethics is grounded in normative theories such as utilitarianism, deontology, and virtue ethics, each offering different criteria for evaluating the moral implications of automated decisions. For instance, utilitarian approaches emphasize outcomes and overall societal benefit, while deontological frameworks focus on adherence to rules and the protection of individual rights. In practice, however, translating these abstract ethical theories into operational guidelines for AI systems remains a significant challenge, particularly given the technical complexity and context-dependent nature of algorithmic decision-making (Mittelstadt et al., 2016).

Another central aspect of AI ethics is addressing the tension between automation and human intervention. As AI systems increasingly take on roles traditionally performed by humans, questions arise about the extent to which decision-making authority should be delegated to machines and how to effectively maintain human oversight. Researchers argue that maintaining effective human oversight is essential to ensuring accountability and preventing unintended harm (Amoroso & Tamburrini, 2020). This perspective highlights that ethical AI is not merely about system performance but also about preserving human autonomy and ensuring that individuals retain the ability to understand and contest decisions that affect them. Moreover, AI ethics extends beyond individual systems to encompass broader societal and institutional dimensions. It considers how AI technologies influence power structures, economic distribution, and social justice. For example, the deployment of AI in areas such as hiring, law enforcement, and credit scoring can have far-reaching implications for equality and inclusion, particularly if existing biases are embedded in algorithmic processes (Varona & Suárez, 2022). Consequently, AI ethics emphasizes the need for inclusive design practices, stakeholder engagement, and continuous monitoring to mitigate risks and promote equitable outcomes. This broader perspective reflects a shift from a purely technical understanding of AI toward a socio-technical framework that recognizes the interplay between technology and society.

Despite the growing consensus on the importance of ethical principles, there is still considerable debate about their practical application. One of the key challenges is the lack of universally accepted standards and the difficulty of operationalizing abstract concepts such as fairness and transparency in diverse contexts. Additionally, ethical guidelines are often voluntary and may lack enforcement mechanisms, limiting their effectiveness in practice (Renda, 2019). This has led to calls for stronger regulatory frameworks and institutional governance structures that can translate ethical principles into binding requirements while maintaining flexibility to accommodate technological innovation. In this context, AI ethics can be understood not as a fixed set of rules but as an evolving framework that must continuously adapt to technological advancements and societal expectations. It requires an ongoing dialogue among developers, policymakers, researchers, and the public to ensure that AI systems are developed and deployed in a manner that maximizes benefits while minimizing harm. Ultimately, the concept of AI ethics underscores the need to balance innovation with responsibility, recognizing that the long-term success and legitimacy of AI technologies depend on their alignment with ethical and human-centered values.

### **3. Ethical Foundations of Artificial Intelligence**

The ethical foundations of AI are grounded in a set of core principles that aim to ensure the responsible development and deployment of intelligent systems. Among these principles, fairness, accountability, transparency, and privacy are widely recognized as essential pillars of ethical AI (Floridi et al., 2018; Leslie, 2019). These dimensions are interconnected and collectively shape how AI systems interact with individuals and society. As AI technologies become increasingly embedded in decision-making processes, the need to operationalize these ethical principles within technical and institutional frameworks has become more urgent. However, implementing these values in practice often involves trade-offs between ethical compliance and system performance, highlighting the complexity of ethical AI governance.

### **3.1 Fairness in AI Systems**

Fairness in AI systems refers to the principle that automated decisions should not result in unjust or discriminatory outcomes across individuals or groups (Varona & Suárez, 2022; Cossette-Lefebvre & Maclure, 2023). Bias in AI can emerge from multiple sources, including imbalanced datasets, historical discrimination embedded in training data, or inappropriate feature selection (Min, 2023; Fletcher et al., 2021). Consequently, AI systems may unintentionally perpetuate or even amplify existing social inequalities, affecting decisions related to employment, healthcare, and financial services. To address these challenges, researchers have developed various fairness-aware techniques, including bias detection tools, algorithmic adjustments, and equitable dataset design (Pagano et al., 2023). However, enforcing fairness constraints often requires modifying model objectives or introducing additional constraints during training, which can negatively impact model accuracy and generalization capabilities (Le Quy et al., 2022). This creates a fundamental tension between fairness and performance, raising important questions about how to prioritize ethical considerations in AI system design.

### **3.2 Accountability and Transparency**

Accountability in AI systems concerns the identification of responsibility for decisions made or influenced by automated models (Busuioac, 2021). Unlike traditional software systems that operate based on explicit and deterministic rules, modern AI systems (particularly those based on deep learning) often function as “black boxes,” making it difficult to trace how specific decisions are generated (Pei et al., 2017). This opacity presents significant challenges for assigning responsibility and ensuring ethical oversight. Transparency and explainability are therefore critical components of accountability, enabling stakeholders to understand, evaluate, and potentially contest AI-driven decisions (Chinnaraju, 2025). The field of explainable artificial intelligence (XAI) has emerged to address these issues by developing methods that provide insights into model behaviour and decision logic (Dwivedi et al., 2023; Das & Rad, 2020). However, current methods of interpreting matters may lead to additional computational complexity and may not fully resolve accountability concerns, particularly in high-stakes contexts where clear and actionable explanations are required.

### **3.3 Privacy and Data Protection**

Privacy and data protection represent fundamental ethical concerns in AI, given the reliance of these systems on large volumes of data for training and decision-making (Duan et al., 2019). This data often includes sensitive personal information such as health records, financial data, and behavioural patterns, raising significant risks related to unauthorized access, misuse, and surveillance (Radanliev et al., 2024). The collection and processing of such data without informed consent or clear purpose limitation can lead to violations of individual rights and erosion of public trust. Moreover, complex data infrastructures are vulnerable to security breaches, which can have serious legal, social, and psychological consequences for affected individuals (Alabi & Holmes, 2024). The commercialization of personal data and its potential transfer to third parties further exacerbate these concerns. As a result, privacy and data protection are considered central pillars of ethical AI, requiring adherence to principles such as data minimization, transparency, user consent, and robust cybersecurity measures (Radanliev et al., 2024). Regulatory frameworks like the GDPR reinforce these principles, although their implementation may introduce challenges for data availability and system performance.

## **4. Current Ethical Concerns Regarding Artificial Intelligence**

The rapid expansion of AI technologies has intensified contemporary ethical challenges, making them among the most pressing issues in the digital era (Bayan & Fayyad, 2024; Dhirani et al., 2023). One major concern is the misuse of AI in surveillance systems, including facial recognition and behavioral monitoring technologies, which are sometimes deployed without adequate safeguards, leading to privacy violations and restrictions on individual freedoms. These technologies can create a “chilling effect,” discouraging free expression and undermining democratic values. Additionally, the increasing reliance on automated decision-making in critical domains such as healthcare, employment, and criminal justice raises concerns about errors,

lack of transparency, and limited avenues for contesting decisions (Helbing, 2018). The ethical implications become even more complex in military applications, particularly with the development of autonomous weapon systems capable of making life-and-death decisions without direct human control (Amoroso & Tamburrini, 2020; Asaro, 2020). These developments challenge traditional notions of accountability and legal responsibility (Schwarz, 2021). Furthermore, AI contributes to broader societal challenges, including the widening digital divide and labor market disruptions. The unequal distribution of AI technologies and skills may exacerbate global inequalities, while automation threatens traditional employment structures (Yu, 2020; Božić, 2023). Without appropriate policy interventions, these trends may deepen social and economic disparities (Kitsara, 2022). These concerns are compounded by the gap between rapid technological advancement and slower regulatory responses, highlighting the need for coordinated international efforts to establish adaptive and ethical governance frameworks.

## 5. Future Trends in AI Ethics

Future developments in AI ethics are increasingly oriented toward the concept of human-centered AI, which emphasizes aligning technological systems with human values, rights, and societal needs. This approach seeks to integrate ethical considerations into all stages of AI development, from design to deployment, ensuring that technology serves humanity rather than undermining it (Floridi et al., 2018). At the global level, there is growing momentum toward the harmonization of ethical standards and regulatory frameworks. International organizations, governments, and academic institutions are collaborating to establish common principles that promote transparency, fairness, and accountability in AI systems. These efforts also emphasize the importance of interdisciplinary dialogue and stakeholder engagement in shaping ethical AI governance (Buhmann & Fieseler, 2021). Moreover, educational institutions are expected to play a crucial role in advancing AI ethics by incorporating ethical training into technical curricula and fostering awareness among future developers and policymakers. As AI continues to evolve, the integration of ethical considerations into research, education, and policy will be essential for ensuring sustainable and responsible innovation. Ultimately, the future of AI ethics lies in its ability to adapt to emerging challenges while maintaining a strong commitment to human dignity and societal well-being.

## 6. Discussion

The findings from the literature and ethical analysis highlight the complex and multifaceted nature of AI ethics. AI systems operate at the intersection of technology, society, and law, and ethical considerations must be embedded across all these dimensions (Floridi et al., 2018; Leslie, 2019). The discussion of fairness, accountability, transparency, and privacy demonstrates that ethical principles cannot be considered in isolation; rather, they interact dynamically and often involve trade-offs that require careful deliberation. For instance, implementing fairness constraints in machine learning models may reduce predictive accuracy, while strict privacy regulations such as the GDPR may limit the availability of data necessary for system optimization (Le Quy et al., 2022; Alic, 2021). The current ethical landscape also underscores the challenges of operationalizing abstract principles in real-world applications. AI technologies are often opaque, highly adaptive, and context-dependent, which complicates the enforcement of accountability and transparency (Busuioc, 2021; Chinnaraju, 2025). Explainable AI methods offer promising approaches, but they are not yet sufficient to guarantee full interpretability, particularly in high-stakes domains such as healthcare, criminal justice, and autonomous weapons systems (Dwivedi et al., 2023; Amoroso & Tamburrini, 2020). Moreover, the global disparity in AI adoption and expertise creates additional ethical concerns regarding the digital divide and social inequality (Yu, 2020; Božić, 2023). Another critical insight from the discussion is the role of governance frameworks in bridging ethical theory and practical implementation. Legal regulations, industry standards, and institutional oversight are crucial for promoting responsible AI while safeguarding human rights (Mitrou, 2018; Sartor & Lagioia, 2020). However, regulatory frameworks must remain flexible and adaptive to avoid stifling innovation or creating unintended consequences. This highlights the importance of collaborative, multi-stakeholder approaches, involving technology developers, policymakers, academia, and civil society, to co-create governance structures that are both effective and

ethically robust (Buhmann & Fieseler, 2021; Herrmann, 2023). Ultimately, the discussion emphasizes that ethical AI is not a purely technical problem but a socio-technical challenge requiring continuous evaluation, reflection, and adaptation. Aligning AI development with human values, legal norms, and societal expectations is essential for ensuring trust, legitimacy, and sustainable benefits, while mitigating risks of harm, bias, or social inequities (Floridi et al., 2018; Renda, 2019).

## 7. Conclusion

The ethical dimensions of AI are central to ensuring that technological advancement benefits society while minimizing harm. AI systems, particularly those deployed in high-stakes domains such as healthcare, finance, and public administration, must operate under principles of fairness, accountability, transparency, and privacy. Ethical adherence is not merely a moral consideration; it is a necessary condition for trust, social acceptance, and sustainable adoption of AI technologies. Implementing ethical AI involves navigating trade-offs between technical performance and adherence to ethical norms. Enforcing fairness constraints or complying with data protection regulations may limit model efficiency or data availability, yet these trade-offs are essential to protect human rights and promote social equity. Transparency and interpretability are also critical for accountability, especially in contexts where automated decisions carry significant consequences for individuals or communities. Furthermore, the responsible development and deployment of AI require coordinated governance, legal frameworks, and institutional oversight. Collaboration among policymakers, researchers, civil society, and technology developers is vital to create flexible regulations and practical ethical guidelines that balance innovation with societal well-being. Ultimately, the future of AI is not only a technological challenge but also a socio-ethical endeavor. Human-centric AI, guided by ethical principles and legal norms, is essential to ensure that these technologies enhance human dignity, support sustainable development, and mitigate risks such as bias, privacy violations, and social inequities. By embedding ethics into AI research, education, policy, and practice, society can harness the transformative potential of AI while safeguarding against its unintended harms.

## References

- Alabi, M., & Holmes, T. (2024). *AI safety and ethics: Developing robust frameworks for ethical AI development and deployment*. ResearchGate.
- Alic, D. (2021). The role of data protection and cybersecurity regulations in artificial intelligence global governance: A comparative analysis of the European Union, the United States, and China regulatory framework. *Search In*.
- Amoroso, D., & Tamburrini, G. (2020). Autonomous weapons systems and meaningful human control: Ethical and legal issues. *Current Robotics Reports*, 1(4), 187–194. <https://doi.org/10.1007/s43154-020-00028-0>
- Asaro, P. (2020). Autonomous weapons and the ethics of artificial intelligence. In *Ethics of artificial intelligence* (pp. 212–236).
- Božić, V. (2023). Artificial intelligence as the reason and the solution of digital divide. *Language Education and Technology*, 3(2).
- Buhmann, A., & Fieseler, C. (2021). Towards a deliberative framework for responsible innovation in artificial intelligence. *Technology in Society*, 64, 101475. <https://doi.org/10.1016/j.techsoc.2020.101475>
- Bayan, F. M., & Fayyad, M. (2024). The ethics of AI: Navigating the moral dilemmas of artificial intelligence. *Arab Journal for Scientific Publishing (AJSP)*, 2663, 5798.
- Busuioac, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, 81(5), 825–836. <https://doi.org/10.1111/puar.13340>
- Chinnaraju, A. (2025). Explainable AI (XAI) for trustworthy and transparent decision-making: A theoretical framework for AI interpretability. *World Journal of Advanced Engineering Technology and Sciences*, 14(3), 170–207.
- Cossette-Lefebvre, H., & Maclure, J. (2023). AI's fairness problem: Understanding wrongful discrimination in the context of automated decision-making. *AI and Ethics*, 3(4), 1255–1269. <https://doi.org/10.1007/s43681-023-00267-5>
- Das, A., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (XAI): A survey. *arXiv preprint arXiv:2006.11371*. <https://arxiv.org/abs/2006.11371>
- Dhirani, L. L., Mukhtiar, N., Chowdhry, B. S., & Newe, T. (2023). Ethical dilemmas and privacy issues in emerging technologies: A review. *Sensors*, 23(3), 1151. <https://doi.org/10.3390/s23031151>
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of big data:

- Evolution, challenges, and research agenda. *International Journal of Information Management*, 48, 63–71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
- 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. <https://doi.org/10.1145/3571732>
- Fletcher, R. R., Nakeshimana, A., & Olubeko, O. (2021). Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health. *Frontiers in Artificial Intelligence*, 3, 561802. <https://doi.org/10.3389/frai.2020.561802>
- Floridi, L. (2023). The ethics of artificial intelligence: Principles, challenges, and opportunities.
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Green, B., & Chen, Y. (2021). Algorithmic risk assessments can alter human decision-making processes in high-stakes government contexts. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–33. <https://doi.org/10.1145/3449181>
- Helbing, D. (2018). Societal, economic, ethical, and legal challenges of the digital revolution: From big data to deep learning, artificial intelligence, and manipulative technologies. In *Towards digital enlightenment: Essays on the dark and light sides of the digital revolution* (pp. 47–72). Springer International Publishing.
- Herrmann, H. (2023). What's next for responsible artificial intelligence: A way forward through responsible innovation. *Heliyon*, 9(3), e14256. <https://doi.org/10.1016/j.heliyon.2023.e14256>
- Kitsara, I. (2022). Artificial intelligence and the digital divide: From an innovation perspective. In *Platforms and artificial intelligence: The next generation of competences* (pp. 245–265). Springer International Publishing.
- Le Quy, T., Roy, A., Iosifidis, V., Zhang, W., & Ntoutsis, E. (2022). A survey on datasets for fairness-aware machine learning. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(3), e1452. <https://doi.org/10.1002/widm.1452>
- Leslie, D. (2019). Understanding artificial intelligence ethics and safety. *arXiv preprint arXiv:1906.05684*. <https://arxiv.org/abs/1906.05684>
- Lombana Diaz, C. (2025). AI ethics. In *Human-Centered AI: An Illustrated Scientific Quest* (pp. 439–474). Springer Nature.
- Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and ‘the dark side’ of AI. *European Journal of Information Systems*, 31(3), 257–268. <https://doi.org/10.1080/0960085X.2021.1986030>
- Min, A. (2023). Artificial intelligence and bias: Challenges, implications, and remedies. *Journal of Social Research*, 2(11).
- Mitrou, L. (2018). Data protection, artificial intelligence and cognitive services: Is the General Data Protection Regulation (GDPR) ‘artificial intelligence-proof’? *Artificial Intelligence and Cognitive Services*.
- Nikolinakos, N. T. (2023). Ethical principles for trustworthy AI. In *EU policy and legal framework for artificial intelligence, robotics and related technologies—the AI Act* (pp. 101–166). Springer International Publishing.
- Pagano, T. P., Loureiro, R. B., Lisboa, F. V., Peixoto, R. M., Guimarães, G. A., Cruz, G. O., ... & Nascimento, E. G. (2023). Bias and unfairness in machine learning models: A systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. *Big Data and Cognitive Computing*, 7(1), 15. <https://doi.org/10.3390/bdcc7010015>
- Patil, D. (2024). Ethical challenges in industrial artificial intelligence applications: Bias, privacy, and accountability. *Privacy, And Accountability* (November 18, 2024).
- Radanliev, P., Santos, O., Brandon-Jones, A., & Joinson, A. (2024). Ethics and responsible AI deployment. *Frontiers in Artificial Intelligence*, 7, 1377011. <https://doi.org/10.3389/frai.2024.1377011>
- Renda, A. (2019). Artificial intelligence: Ethics, governance, and policy challenges. CEPS Centre for European Policy Studies.
- Sartor, G., & Lagoia, F. (2020). The impact of the General Data Protection Regulation (GDPR) on artificial intelligence.
- Varona, D., & Suárez, J. L. (2022). Discrimination, bias, fairness, and trustworthy AI. *Applied Sciences*, 12(12), 5826. <https://doi.org/10.3390/app12125826>
- Yu, P. K. (2020). The algorithmic divide and equality in the age of artificial intelligence. *Florida Law Review*, 72, 331–374.