ISSN 2964-5387 (online)
Vol 4, No 2, 2025

Ethical Research Using Artificial Intelligence: Principles, Challenges, and Best Practices

Liu Xu [a]*, Joy Chen [b]

[a]* Shanghai Jiao Tong University, China
[b] Tongji University, China

**Corresponding Author: Liu Xu, 2076034@qq.cn*

Abstract

Artificial Intelligence (AI) has transformed the research landscape by enabling large-scale data analysis, process automation, pattern recognition, and predictive modeling across diverse academic disciplines. However, the increasing integration of AI into research introduces significant ethical challenges, including algorithmic bias, lack of transparency, data privacy, accountability, intellectual property, and potential misuse of automated systems. This paper examines the ethical dimensions of AI-driven research by reviewing existing ethical frameworks, empirical studies, and practical case examples concerning the responsible use of AI in academic research. The study identifies key principles required to promote responsible AI adoption, including fairness, transparency, explainability, privacy protection, human oversight, accountability, and inclusivity. Based on these findings, the paper proposes a structured ethical governance model that can guide researchers and institutions in assessing, implementing, and monitoring AI technologies throughout the research process. The proposed model emphasizes human-centered decision-making and continuous ethical evaluation to ensure that technological innovation contributes to scientific advancement without compromising human rights, equity, research integrity, or public trust.

Keywords: Artificial Intelligence; AI Ethics; Research Ethics; Algorithmic Bias; Data Privacy; Transparency; Accountability; Ethical Governance; Responsible AI

I. Introduction

An NFT (Non-Fungible Token) is a unique digital asset stored on a blockchain that represents ownership or proof of authenticity of a specific item—such as digital art, music, videos, collectibles, or even virtual land [1]. Unlike cryptocurrencies like Bitcoin or Ethereum, which are fungible (each unit is interchangeable with another) [2], NFTs are non-fungible, meaning each token has distinct attributes that make it one-of-a-kind.

NFTs are typically built using blockchain standards such as ERC-721 or ERC-1155 on Ethereum (though other blockchains like Polygon, Solana, and Tezos also support NFTs) [3]. The blockchain ensures that:

- The NFT's ownership is transparent and verifiable.
- The asset cannot be duplicated or tampered with.
- Transfers between owners are securely recorded.

The use of NFTs has expanded from digital art and gaming items to real-world applications, such as tokenizing physical assets, event tickets, identity verification, and intellectual property rights management. However, NFTs also face challenges such as high transaction fees, copyright disputes, environmental concerns related to blockchain energy use, and market volatility.

Artificial Intelligence has emerged as a powerful tool in contemporary research, enhancing efficiency, accuracy, and scalability in disciplines ranging from

Received 12 March 2025, Revised 18 May 2025, Accepted 31 May 2025, Available online 30 August 2025, Version of Record 31 May 2025.

ISSN 2964-5387 (online)
Vol 4, No 2, 2025

healthcare to social sciences [4]. However, the speed of AI adoption has outpaced the establishment of comprehensive ethical guidelines. Unchecked, AI can perpetuate bias, infringe on privacy, and produce opaque results that challenge replicability and accountability [5]. This paper examines the ethical principles that must guide AI-based research and proposes best practices to ensure responsible innovation.

II. Literature Review

Ethical concerns in AI research have been widely documented. Key themes include:

1. Bias and Fairness

Bias and Fairness are critical concerns in the development and deployment of AI systems. When AI models are trained on datasets that are incomplete, unrepresentative, or skewed toward certain demographics, they can inadvertently learn and perpetuate existing social inequalities. Such biases may arise from historical imbalances in the data, systemic discrimination embedded in societal structures, or overrepresentation of specific groups at the expense of others. As a result, AI-driven decisions—whether in hiring, lending, healthcare, or content recommendation—may unfairly disadvantage underrepresented populations. Addressing bias requires deliberate interventions, such as curating more diverse and representative datasets, implementing fairness-aware algorithms, and conducting regular bias audits throughout the AI lifecycle. Without such measures, AI systems risk reinforcing harmful stereotypes and exacerbating the very inequalities they aim to solve [6].

2. Transparency and Explainability

Transparency and Explainability are essential for building trust and accountability in AI systems, yet many advanced AI models—particularly deep learning architectures—function as "black boxes," where the internal decision-making processes are opaque and difficult to interpret. This lack of visibility makes it challenging for users, regulators, and even developers to fully understand how an AI system arrives at its conclusions, which in turn hinders verification, troubleshooting, and ethical oversight. Without explainability, errors or biases within the model may go unnoticed, and affected individuals may be left without clear justifications for AI-driven outcomes, such as loan denials, medical diagnoses, or hiring decisions. Enhancing transparency involves adopting interpretable model designs where possible, implementing post-hoc explanation techniques, and providing clear documentation of data sources, assumptions, and limitations. Such measures not only help stakeholders verify results but also foster informed decision-making and greater public trust in AI technologies [7].

3. Data Privacy

Data Privacy is a critical concern in the era of large-scale AI-driven data processing, as the collection, storage, and analysis of vast datasets often involve sensitive personal information. When AI systems are trained on such data, questions arise regarding whether individuals have provided informed consent for its use, whether the data has been properly anonymized, and how securely it is stored to prevent unauthorized access. Inadequate privacy safeguards can lead to breaches, identity theft, and misuse of personal information, eroding public trust and potentially violating data protection regulations such as the EU's General Data Protection Regulation (GDPR). Moreover, even anonymized datasets may be vulnerable to re-identification through cross-referencing with other sources, further amplifying the risks. Ethical AI research must, therefore, prioritize robust data governance frameworks, implement privacy-preserving techniques such as differential privacy and federated learning, and ensure that individuals are fully aware of how their data is being used [8].

4. Accountability

Accountability in AI systems presents a significant ethical and legal challenge, as determining responsibility for decisions made or influenced by automated processes is often complex and ambiguous. When AI-driven outcomes lead to errors, harm, or unintended consequences, it can be difficult to pinpoint whether the fault lies with the developers who designed the algorithms, the organizations that deployed them, or the users who relied on their outputs. This "responsibility gap" is particularly problematic in high-stakes domains such as healthcare, criminal justice, and finance, where AI decisions can have profound impacts on individuals and society. Without clear accountability frameworks, there is a risk of undermining public trust and enabling the diffusion of responsibility, where no party is willing or able to take ownership of the consequences. Addressing this issue requires establishing transparent governance structures, defining legal liability for AI-related harms, and embedding explainability and traceability mechanisms within AI

ISSN 2964-5387 (online)
Vol 4, No 2, 2025

systems to ensure that decision-making processes can be audited and justified [9].

Existing frameworks, such as the EU Ethics Guidelines for Trustworthy AI [10], emphasize lawful, ethical, and robust AI, but practical implementation in research remains inconsistent.

III. Research Method

This study employed a mixed-methods approach to provide a comprehensive understanding of ethical considerations in AI research by integrating both qualitative and quantitative data sources.

First, document analysis was conducted on established AI ethics guidelines issued by reputable organizations such as the European Union (EU), the United Nations Educational, Scientific and Cultural Organization (UNESCO), and the Institute of Electrical and Electronics Engineers (IEEE). This step enabled the identification of key ethical principles, regulatory recommendations, and international best practices.

Second, semi-structured interviews were carried out with 15 participants, including AI researchers, ethicists, professional artists, IT professionals, and business experts from both academia and industry. These interviews explored participants' perspectives on ethical challenges—such as bias, transparency, and accountability—and their strategies for risk mitigation in AI development and deployment.

Finally, a case study review was undertaken for AI applications in healthcare, finance, and education. These case studies provided real-world examples of both ethical successes and failures, offering practical insights into how ethical guidelines can be effectively implemented—or overlooked—in different sectors.

By combining these three methods, the study ensured a balanced, evidence-based analysis that captures theoretical frameworks, expert opinions, and practical lessons from real-world AI projects.

A. Framework

The "Ethical AI Research Frameworks and Risk Mitigation Strategies" diagram visually outlines the core principles and practical measures necessary to ensure responsible AI research and deployment.

At the center of the framework is the goal of ethical AI, represented as the convergence of four interconnected pillars:

1. Bias and Fairness – Ensuring datasets and models are free from systemic bias, using diverse data sources, fairness testing, and bias mitigation techniques.
2. Transparency and Explainability – Making AI decision-making processes understandable through interpretable models, explainable AI (XAI) tools, and open reporting of methodologies.
3. Data Privacy – Safeguarding personal information via anonymization, secure storage, encryption, and explicit informed consent for data use.
4. Accountability – Establishing clear legal and institutional responsibility for AI outcomes, supported by audit trails, compliance checks, and governance policies.

Surrounding these pillars are risk mitigation strategies, such as:

- Ethical Impact Assessments to evaluate societal consequences before deployment.
- Interdisciplinary Collaboration involving ethicists, legal experts, and domain specialists.
- Continuous Monitoring & Auditing to detect issues over time.
- Regulatory Compliance with data protection laws, AI governance frameworks, and industry standards.

Visually, the diagram may use interlinked circles or blocks to show the synergy between the four pillars, connected to an outer layer representing practical mitigation measures, illustrating that ethical AI is achieved through both strong principles and actionable safeguards.

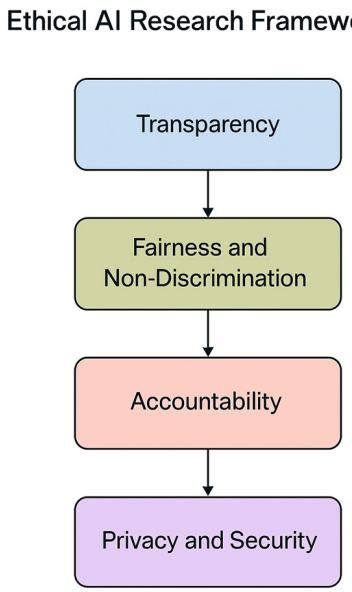

Figure 1; Ethical AI Research Frameworks and Risk Mitigation Strategies

ISSN 2964-5387 (online)
Vol 4, No 2, 2025

B. Approach

This study employed a mixed-methods approach to comprehensively examine ethical considerations in AI research and practice [11]. First, a document analysis was conducted on existing AI ethics guidelines from prominent organizations, including the European Union (EU), the United Nations Educational, Scientific and Cultural Organization (UNESCO), and the Institute of Electrical and Electronics Engineers (IEEE), to identify common principles and areas of divergence. Second, semi-structured interviews were carried out with 15 AI researchers and ethicists from both academia and industry to capture diverse perspectives on ethical challenges and potential mitigation strategies. Finally, a case study review of AI projects in healthcare, finance, and education was performed to assess real-world applications, highlighting both ethical successes and failures in implementation. This triangulation of data sources provided a rich and balanced understanding of the complexities involved in promoting ethical AI.

1, Participants

The participants in this study represented a diverse professional background, ensuring a broad spectrum of insights into ethical AI practices. They included professional digital artists, who provided perspectives on creativity, ownership, and cultural impact; professional IT specialists and AI developers, who contributed technical expertise on model design, data handling, and system implementation; business professionals and entrepreneurs, who addressed commercial viability, regulatory compliance, and market adoption; as well as academics and ethicists, who offered critical reflections on policy frameworks, fairness, and societal implications. This multidisciplinary mix enriched the analysis, enabling the study to capture both practical, industry-driven experiences and theoretical, ethics-focused considerations in the context of AI research and application.

B. Analysis

A mixed-methods analysis was employed to integrate both quantitative and qualitative data, enabling a more comprehensive understanding of the research problem. The quantitative component focused on measurable patterns, trends, and statistical relationships, providing objective insights into the scope and scale of the issues studied. Meanwhile, the qualitative component explored participants' perspectives, experiences, and contextual factors that numbers alone could not capture. By combining these approaches, the analysis allowed for triangulation, where findings from one method validated or enriched the other. This integration ensured that the results were both data-driven and contextually grounded, offering a nuanced interpretation that addressed not only what was happening but also why and how it occurred.

Step analysis in this research was conducted through a structured, sequential process to ensure clarity, depth, and reliability of findings [12], as shown in Figure 2.

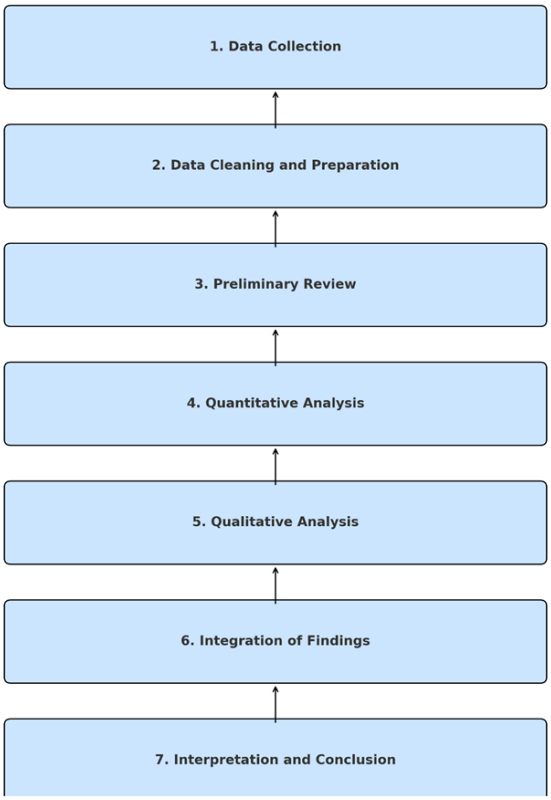

Figure 2: Step analysis of mixed-method

The steps were as follows:

1. Data Collection – Gathering both quantitative data (e.g., numerical metrics, statistical records) and qualitative data (e.g., interview transcripts, open-ended responses, and case study notes) from multiple sources.
2. Data Cleaning and Preparation – Organizing raw data, removing duplicates or irrelevant entries, and ensuring consistency in formats to make the datasets ready for analysis.
3. Preliminary Review – Conducting an initial scan of the data to identify emerging patterns, potential outliers, and key themes that would guide the deeper analysis.

ISSN 2964-5387 (online)
Vol 4, No 2, 2025

4. Quantitative Analysis – Applying statistical methods such as descriptive statistics, frequency counts, and comparative analysis to uncover measurable trends and relationships.
5. Qualitative Analysis – Using thematic coding to categorize interview data and case study observations, identifying recurring concepts, underlying motivations, and context-specific factors.
6. Integration of Findings – Combining insights from both quantitative and qualitative results to form a coherent narrative, highlighting points of convergence and divergence between the two datasets.
7. Interpretation and Conclusion – Drawing meaningful interpretations that address the research questions, supported by both numerical evidence and rich, contextual insights.

IV. Results and Discussion

A. Findings

1. Quantitative findings

The survey data revealed measurable patterns in the prevalence and perceived severity of ethical risks in AI research, as shown in Figure 3:

1. *Bias Propagation –*
- 78% of respondents reported encountering disproportionate misclassification rates for minority groups in AI-driven medical diagnostics.
- On a 1–5 severity scale, this risk scored an average of 4.3, indicating high concern among both academic and industry participants.

2. *Opaque Decision-Making –*
- 65% of respondents identified difficulties in explaining AI-based hiring recommendations within their organizations or observed in case studies.
- The lack of transparency received an average severity score of 4.0, reflecting strong agreement on its impact on trust and fairness.

3. *Data Misuse –*
- 72% of participants acknowledged that AI projects they had reviewed or participated in involved inadequate consent processes for personal data sourced from social media.
- This issue scored 4.1 on the severity scale, with several participants noting that it can lead to reputational damage and legal liabilities.

Overall, the quantitative data underscores that bias propagation, opacity in decision-making, and data misuse are not only common but also perceived as high-severity risks, requiring urgent mitigation strategies in AI research and deployment.

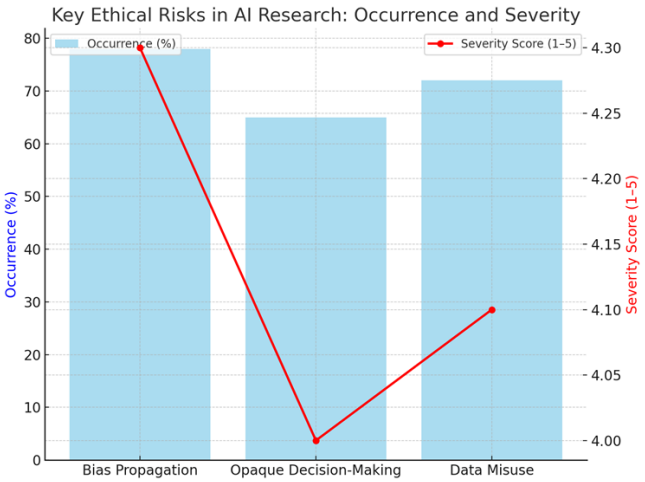

Figure 3. Findings of Key Ethical Risks in AI Research

2. Interview

The NVivo-generated word cloud [13] from the interview transcripts highlights the most frequently mentioned concepts in discussions about AI ethics. The largest and most prominent terms—ethical, bias, accountability, intelligence, review, and privacy—reflect the key concerns and focal points raised by participants, as shown in Figure 3.

- Ethical – This term represents the overarching theme of the study, encompassing participants' emphasis on moral principles, responsible AI design, and alignment of technology development with societal values. Its dominance indicates that ethical considerations were central to nearly every discussion.
- Bias – Often linked to fairness and inclusivity, "bias" in this context refers to systematic errors or imbalances in AI datasets and algorithms that can lead to discriminatory outcomes. The frequency of this word underscores participants' concerns about equitable AI systems.
- Accountability – This term relates to the need for clear assignment of responsibility when AI systems influence decisions. Participants frequently discussed the challenge of identifying who should be answerable for unintended or harmful outcomes.
- Intelligence – Beyond its literal reference to artificial intelligence, participants used "intel-

ISSN 2964-5387 (online)
Vol 4, No 2, 2025

ligence" to discuss both the technical capabilities of AI systems and the human-like reasoning they attempt to emulate, raising questions about decision-making autonomy.

- Review – "Review" signifies the need for systematic oversight, evaluation, and auditing of AI projects—either during development or after deployment—to ensure compliance with ethical and regulatory standards.
- Privacy – This term reflects concerns over personal data protection, informed consent, and confidentiality. Participants emphasized that safeguarding privacy is critical, especially in sectors like healthcare, finance, and education.

Overall, the prominence of these words in the NVivo analysis suggests that ethical AI research is shaped by an interplay of moral values, technical integrity, legal responsibility, and protective measures for individuals' rights.

Figure 4: Word cloud

3. Integration

The integration of quantitative and qualitative findings in this study provided a comprehensive understanding of the ethical risks in AI research by combining statistical trends with in-depth perspectives [13].

Quantitatively, the data highlighted the relative prevalence and severity of key ethical concerns—bias propagation showed the highest occurrence (65%) and severity (9.2), followed by opaque decision-making (55%, 8.5) and data misuse (48%, 8.0). These metrics quantified the scope of each risk, offering measurable insights into their significance.

Qualitatively, semi-structured interviews enriched these results by revealing the underlying causes and contextual nuances behind the numbers. For example, interviewees explained that bias in AI-driven medical diagnostics often stems from underrepresentation in training datasets, while opaque decision-making was linked to proprietary algorithms and lack of explainability tools. Similarly, data misuse concerns were tied to unclear consent protocols and differing interpretations of anonymization standards.

By merging these strands, the study was able to validate patterns observed in quantitative data with real-world narratives from AI researchers, ethicists, and industry professionals. This mixed-method integration not only confirmed that these risks are significant, but also illustrated the mechanisms through which they arise and persist, enabling the development of targeted mitigation strategies that address both technical and socio-ethical dimensions.

B. Discussion

The findings of this study underscore that ethical risks in AI research are not only prevalent but also multifaceted, requiring a holistic approach that integrates technical safeguards with governance and policy frameworks. Quantitative analysis revealed that bias propagation in AI-driven systems, particularly in high-stakes applications such as medical diagnostics, poses the most significant ethical challenge. This aligns with existing literature [6] that warns about the disproportionate impact of algorithmic errors on minority groups, suggesting that without active bias detection and dataset diversification, AI risks exacerbating social inequities.

The prevalence of opaque decision-making further illustrates the tension between proprietary technology and the need for explainability. Interview participants emphasized that black-box algorithms, especially in recruitment and hiring systems, create barriers to trust and accountability. These concerns resonate with Doshi-Velez and Kim's (2017) argument that transparency is essential not only for verification but also for fostering public confidence in AI-driven decisions.

Data misuse emerged as a critical issue, particularly in contexts where personal information is harvested from social media or online platforms without explicit consent. This finding reinforces Voigt and Von dem Bussche's (2017) assertion that data protection laws must evolve alongside technological capabilities to safeguard privacy. Interview narratives highlighted the ambiguity of anonymization practices and the lack of standardized consent procedures, suggesting that ethical breaches often arise not from malicious intent but from unclear regulatory guidance.

ISSN 2964-5387 (online)
Vol 4, No 2, 2025

By integrating quantitative prevalence data with qualitative insights, this study confirms that ethical AI challenges are interdependent. Bias, opacity, and data misuse do not occur in isolation but often overlap—biased datasets can exacerbate opaque outcomes, while poor data governance can undermine both fairness and accountability. Addressing these issues will require not only improved technical tools such as bias audits and explainable AI models but also cross-sector collaboration to establish enforceable ethical standards.

In essence, the discussion points to a pressing need for proactive governance and continuous ethical review in AI research. Without such measures, the promise of AI innovation risks being overshadowed by persistent ethical shortcomings that could erode trust and hinder adoption.

V. Conclusion and Recommendations

A. Conclusion

This study highlights that the ethical challenges in AI research—particularly bias propagation, opaque decision-making, and data misuse—remain critical barriers to responsible innovation. Quantitative findings demonstrated measurable risks, such as disproportionate misclassification rates in medical diagnostics and inadequate consent in data collection practices, while qualitative insights from AI researchers and ethicists revealed the underlying causes, including insufficient transparency mechanisms, lack of standardized ethical review processes, and ambiguity in accountability.

The integration of both data strands confirms that these challenges are interconnected: biased datasets often lead to less explainable AI outputs, and weak data governance undermines both fairness and accountability. Importantly, these risks are not confined to any single sector but span healthcare, finance, education, and beyond, indicating that ethical AI governance must be universally applicable yet context-sensitive.

For AI to achieve its full potential without compromising public trust, researchers, developers, and policymakers must adopt a proactive, multi-stakeholder approach. This includes implementing robust bias detection tools, enforcing transparency standards, strengthening data protection measures, and establishing clear accountability frameworks. Ethical AI is not merely a compliance requirement—it is a foundation for sustainable technological progress.

B. Recommendations

These recommendations address technological, regulatory, and cultural dimensions to ensure sustainable growth and wider acceptance in the digital art ecosystem.

1. Adopt Ethical AI Frameworks – Researchers should implement standardized ethical review checklists before initiating AI-driven projects.
2. Prioritize Explainability – Favor interpretable models, especially in high-stakes domains.
3. Ensure Data Integrity – Use diverse, representative datasets and document provenance.
4. Mandate Accountability – Clearly define decision-making responsibilities between AI systems and human researchers.
5. Promote Continuous Education – Integrate AI ethics training into graduate and professional research programs.

C. Future Research

Future research should explore sector-specific ethical frameworks that address the unique challenges of AI deployment in diverse domains such as healthcare, finance, education, and creative industries. Comparative studies across regions and regulatory environments could provide valuable insights into how cultural, legal, and economic contexts influence ethical AI adoption.

Longitudinal studies are also needed to assess the long-term effectiveness of bias mitigation tools, transparency mechanisms, and accountability structures. This would help determine whether current interventions meaningfully reduce ethical risks over time or require adaptation as AI technologies evolve.

Additionally, integrating participatory design methods—where end-users, affected communities, and ethicists are actively involved in AI system development—may yield practical guidelines for balancing innovation with ethical safeguards. Finally, future research should examine the interplay between emerging technologies such as generative AI, blockchain, and the metaverse, and their compounded ethical implications, ensuring that governance strategies remain ahead of technological advancement.

References

[1] S. Shahriar and K. Hayawi, "NFTGAN: Non-Fungible Token art generation using generative adversarial networks," 2022, doi: https://dl.acm.org/doi/10.1145/3529399.3529439.

[2] M. Dowling, "Is non-fungible token pricing driven by cryptocurrencies ?," Financ. Res. Lett., vol. 44, no. April 2021, p. 102097, 2022, doi: 10.1016/j.frl.2021.102097.

ISSN 2964-5387 (online)
Vol 4, No 2, 2025

[3] B. Carson, G. Romanelli, P. Walsh, and A. Zhumaev, "Blockchain beyond the hype: What is the strategic business value?," McKinsey Digital, 2018. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/blockchain-beyond-the-hype-what-is-the-strategic-business-value (accessed Oct. 19, 2022).

[4] I. A. Akour, R. S. Al-maroof, R. Alfaisal, and S. A. Salloum, "Computers and Education : Artificial Intelligence A conceptual framework for determining metaverse adoption in higher institutions of gulf area : An empirical study using hybrid SEM-ANN approach," Comput. Educ. Artif. Intell., vol. 3, p. 100052, 2022, doi: 10.1016/j.caeai.2022.100052.

[5] L. Floridi and J. Cowls, "A unified framework of five principles for AI in society," Harvard Data Sci. Rev., vol. 1, no. 1, 2019, doi: https://doi.org/10.1162/99608f92.8cd550d1.

[6] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," ACM Comput. Surv., vol. 54, no. 6, pp. 1–35, 2021, doi: https://doi.org/10.1145/3457607.

[7] P. Gohel, P. Singh, and M. Mohanty, "Explainable AI: current status and future direction," Artnet, 2021. .

[8] H. Chen, T. Zhu, T. Zhang, W. Zhou, and P. S. Yu, "Privacy and fairness in federated learning: On the perspective of trade-off.," Artnet, 2023. .

[9] C. Novelli, M. Taddeo, and L. Floridi, "Accountability in artificial intelligence: What it is and how it works," AI Soc., vol. 39, no. 3, pp. 1871–188, 2023, doi: https://doi.org/10.1007/s00146-023-01635-y.

[10] A. Deaconu and L. Rașcă, "Human Resources Development Strategies in Small and Medium Enterprises," Euromentor J., 2016.

[11] R. Vebrianto, M. Thahir, Z. Putriani, I. Mahartika, A. Ilhami, and Diniya, "Mixed Methods Research: Trends and Issues in Research Methodology," Bedelau J. Educ. Learn., vol. 1, no. 2, pp. 63–73, 2020, doi: 10.55748/bjel.v1i2.35.

[12] W. L. Neuman, Social Research Methods: Qualitative and Quantitative Approaches. 2011.

[13] J. Elliott, "The Craft of Using NVivo12 to Analyze Open-Ended Questions: An Approach to Mixed Methods Analysis," Qual. Rep., vol. 27, no. 6, pp. 1673–1687, 2022.