7th European Conference on Industrial Engineering and Operations Management
Augsburg, Germany, July 16-18, 2024
Publisher: IEOM Society International, USA DOI: 10.46254/EU07.20240168
Published: July 16, 2024

Ethical AI Frameworks: Balancing Privacy, Consent and Responsible Use

Tawanda Kapuya, Walter Kubiku, Mqamlandaba Dube, Tafadzwa Mukudu, Mind Kutyauripo, Belinda Ndlovu and Sibusisiwe Dube
Department of Informatics and Analytics
National University of Science and Technology
Bulawayo, Zimbabwe
tlkapuya@gmail.com; walterkubiku@gmail.com; morthansecurity@gmail.com; tafadzwamukudu@gmail.com; mindkutya@gmail.com; belinda.ndlovu@nust.ac.zw ; sibusisiwe.dube@nust.ac.zw

Abstract

The broad adoption of artificial intelligence (AI) technologies across industries has raised several important ethical debates. Concerns that AI systems will exhibit biases, affect individual privacy, and raise issues of accountability surrounding the use of advanced technologies are some of the most critical points. While most AI frameworks address responsible use, the importance of informed consent and data privacy are some grey areas that remain untouched by most researchers due to the dynamic nature of AI development itself. This paper aims to develop a holistic ethical AI framework that balances ethics, privacy, consent, and responsible use. The researchers conducted a systematic literature review (SLR) using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) spanning the years 2020 to 2024, with IEEE Xplore, Web of Science, PubMed, and ACM being the primary databases. A total of 29 papers met the final inclusion criteria. The researchers managed to identify the main features of AI frameworks as privacy, transparency, fairness, trust, security, and accountability. Other features were responsible use, good governance, autonomy, data protection, and explainability. The tools used to evaluate existing ethical frameworks in the observed studies were Bias Detection Algorithms, Deep Learning techniques, and Natural Language Processing. Most studies did not include any tools for evaluating ethical frameworks. According to the observed studies, healthcare was the most prominent application area contributing to more than half compared to other areas like finance, education, transport, information technology, and governance. The main challenges surrounding the privacy and security of AI frameworks were a lack of transparency and compliance, raising concerns about data misuse and user privacy. Security vulnerabilities and breaches of user information further highlight the need for stricter governance and user control. This paper will provide a comprehensive analysis of the research findings and identify research gaps for future research such as aspects/features, application areas, and tools used to evaluate the ethical soundness and challenges surrounding the privacy and security of AI frameworks.

Keywords
Artificial Intelligence (AI), Frameworks, Privacy, Consent, Ethics, Responsible Use.

1. Introduction
Artificial intelligence (AI) is rapidly transforming our world, with applications impacting nearly every facet of society. AI is the new gold rush of the 21st century with tech giants like Meta, Open AI, and Google scrambling to have a superior model that will take a large piece of the market (Landi 2024). This progress, however, is not without its challenges. According to Korobenko et al. (2024), this has caused AI developers to go to extreme lengths to get training data to win the AI race, resulting in many ethical meltdowns and privacy breaches. Concerns regarding the ethical implications of AI development and deployment are growing (Korobenko et al. 2024). At the forefront of these

concerns lie issues of privacy, informed consent, and responsible use (Cumming et al. 2024). The vast amount of data required to train AI models raises concerns about individual privacy and the potential misuse of personal information. Additionally, ensuring users understand and have control over how their data is used is critical for building trust and fostering ethical AI practices. To navigate these complexities, robust ethical frameworks are essential (Eitel-Porter. 2021). However, the landscape of AI frameworks is constantly evolving. Peters et al. (2020) emphasizes the importance of establishing robust ethical frameworks to guide the design and deployment of AI systems as their development accelerates. These frameworks need to consider the complex and rapidly changing landscape of AI capabilities and applications. Cumming et al. (2024) highlights the dynamic nature of the landscape of AI frameworks, with new proposals and refinements emerging regularly. The fluidity in this landscape reflects the dynamism of the field but presents challenges in establishing stable and widely accepted ethical guidelines for AI systems. Reviews in the education sector Cumming et al. (2024); Dube et al. (2023); Oluwaseun Augustine Lottu et al. (2024) have mainly concentrated on the development of AI frameworks focusing on benefits and challenges, whilst Klarin et al. (2024) have focused on how to instil ethicality into AI technologies. In the health sector reviews Korobenko et al. (2024) have mainly focused on establishing a unified definition of key concepts in AI ethics. Despite their relevance, these frameworks have not included ethical aspects such as holistic integration so as to weave the features together into a cohesive framework. Furthermore, with the advent of online learning since the Covid-19 pandemic, studies allude that there are security, privacy, and ethical concerns that still need to be addressed in the educational sector to fully harness online teaching capabilities (Dube et al. 2023). A question remains on how we can make these frameworks ethical. It is against this background that this research seeks to develop a holistic AI framework. To achieve this, the review will identify and synthesise the core concerns, mechanisms, and strategies that underpin ethical AI frameworks. The research discusses the ethical concerns surrounding the development and deployment of artificial intelligence (AI) systems.

1.1 Objectives

- To identify key features of AI frameworks.
- To identify tools that have been used to evaluate the ethical soundness of AI frameworks.
- To identify the application areas of AI Frameworks.
- To identify challenges surrounding the privacy and security of AI frameworks.

2. Methodology

A systematic literature review (SLR) methodology was employed to answer the research questions. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) steps included identification, screening, and eligibility and structured the literature analysis of this study (Moher et al. 2009).

2.1. Database and Search Strategy

The peer-reviewed and published papers analysed in this study were sourced from four scientific databases, namely IEEE Xplore, ACM Digital Library, Web of Science, and PubMed. A search term was constructed to query the above databases and it had some variations to suit the syntax of the database.

The structure of the search term was as follows: ("AI Framework" OR "Artificial Intelligence Framework") AND ("Ethics" OR "Ethical") AND ("Privacy" OR "Security"). Execution of the search query yielded 397 articles distributed as follows: IEEE Xplore n = 84 articles, ACM Digital Library n = 92 articles, PubMed = 102 articles, and Web of Science n = 119 articles.

2.2. Inclusion and exclusion criteria

The systematic literature review's inclusion criteria were limited to peer-reviewed journal articles that were released in the years 2020–2024. The review only included articles published in English. Studies that focused on the use of AI-powered technologies, such as machine learning and artificial intelligence. The frameworks and strategies used to guarantee the moral and responsible use of these technologies were the main interventions of interest. Studies were excluded if they were editorials, letters to the editor, commentaries, or grey literature. Purely descriptive studies without any quantitative or qualitative analysis were also excluded.

2.3. Screening

The initial search of all the databases yielded 397 peer-reviewed articles. After duplicate removal of 28 papers, the identified records were 369, where 286 of them were excluded for the following reasons: not having evidence about ethical AI frameworks, unrelated to data science and AI, or having misalignment with the research questions. After reviewing the full-text articles against the inclusion criteria, 83 studies were further considered for full-text review.

2.5. Included

After a full-paper review of 83 papers, only 29 were included after applying the inclusion criteria. While some papers were prima facie unrelated to the domains of data science and AI, many of these turned out to be highly relevant to the thematic area into which AI ethics, privacy and security intersect.

3. Results

The flow diagram for this study using PRISMA is depicted in Figure 1 and Table 1.

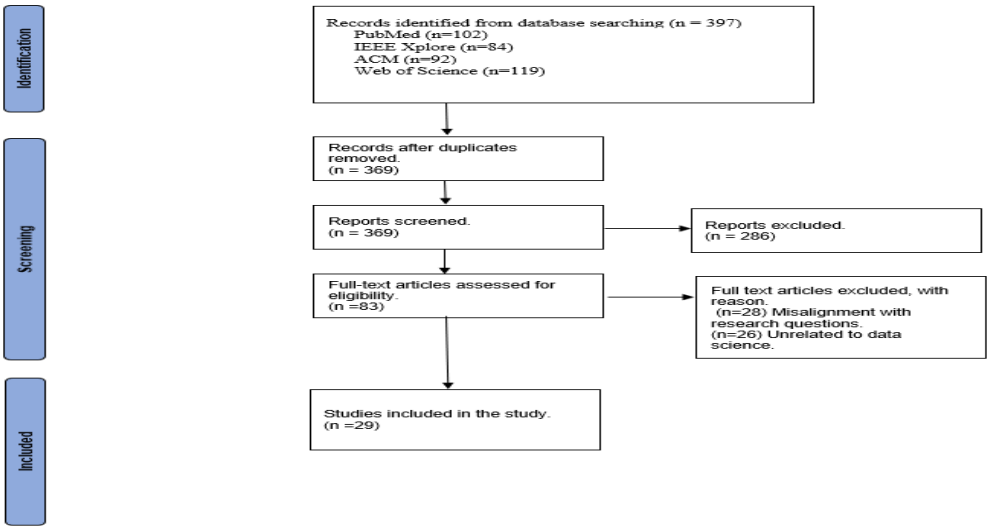

Figure 1. PRISMA Flow Diagram (Moher et al. 2009)

Table 1. Papers that met the inclusion criteria.

Author	Country	Features	Applications areas.	Dimension	Challenges
Oluwabukunmi Latifat Olorunfemi et al. (2024)	USA	● Accountability ● Fairness ● Privacy Protection ● Security ● Fairness	● Health care systems ● Financial systems ● Judiciary systems	● Healthcare ● Finance ● Banking ● Social Media ● Education ● Transport ● Legal and Justice Systems	● Lack of transparency ● ●
Williamson and Prybutok (2024)	USA	● Transparency ● Governance ● Value	● Health care systems ● Legal ● Bias detection	● Health ● ●	● Lack of privacy ● Lack of data integrity ● Lack of AI regulations
Lu et al. (2024)	Australia	● Trustworthiness ● Accountability	● AI Governing bodies	● Governance	● Lack of AI regulations
Kumar et al. (2024)	USA	● Sustainability ●	● Operations Management	● Organisational	● Lack of awareness of AI

		•			
Singhal et al. (2024)	Canada	• Fairness • Transparency • Accountability	• Disease Surveillance •	• Health •	• Lack of security • Lack of privacy
Korobenko et al. (2024)	Estonia	• Security	• Disease Surveillance •	• Health •	• Lack of privacy • Lack of transparency • Lack of accountability
Oluwaseun Augustine Lottu et al. (2024)	UK	• Fairness • Transparency • Accountability • Security	• Academics •	• Education •	• Lack of AI regulations • Lack of privacy • Lack of consent • Lack of security, • Lack of transparency • Lack of accountability
Cumming et al. (2024)	USA	• Autonomy • Biasedness • Explainability	• Academics	• Education	• Lack of data integrity. • Lack of consent • Lack of transparency
Pant et al. (2024)	USA	• Transparency • Privacy	• Image and Speech Recognition • Predictive Analytics	• Health	• Lack of Privacy • Lack of consent. • Lack of security.
Saheb (2024)	USA	• Protection of Personal Data • Security	• AI development at the national government level	• Healthcare • Education • Public Administration • Transport • Security and Defence	• Lack of Privacy • Lack of security • Lack of consent • Lack of Regulatory Compliance
Obreja et al. (2024)	Romania	• Ethical/Moral Dimensions • Balancing Certainties and Uncertainties.	• General AI Innovations	• Sociology • Healthcare • Finance • Education • Transportation	• Lack of Regulatory Compliance •
Rai (2024)	UK	• Privacy • Transparency • Accountability • Public awareness	• Development of an ethical AI framework	• All industries	• Lack of Privacy • Lack of security, • Lack of transparency • Lack of consent
Coiduras-Sanagustín et al. (2024)	Spain	• Privacy	• Product design on personal data privacy in the context of the Internet of Things (IoT).	• Information Technology	• Lack of Privacy
Klarin et al. (2024)	Australia	• Fairness • Responsibility • Trust	• AI development and operation.	• Healthcare • Finance • Banking • Education • Transport	• Lack of Regulatory Compliance
Chen et al. (2023)	Australia	• Trust • Safety • Fairness • Transparency	• Climate change	• Agriculture	• Lack of Regulatory Compliance •
Tahri Sqalli et al. (2023)	Qatar	• Transparency • Fairness • Justice • Accountability	• Medical training • E-health wearables development	• Healthcare •	• Lack of transparency, • Lack of accountability • Lack of collaboration.
Radanliev et al. (2024)	UK	• Fairness • Responsibility • Trust • Privacy	• Privacy-Preserving AI Systems •	• Business Operations	• Lack of Privacy • Lack of security • Lack of transparency • Lack of consent
Schmid and Wiesche (2023)	Germany	• Trust	• Patient management	• Health	• Lack of Privacy • Lack of security,

					● Lack of transparency ● Lack of consent
Prem (2023)	Austria	● Privacy ● Security	● Healthcare ● Medical Diagnosis	● Health	● Lack of transparency.
Li (2023)	China	● Trust ● Security ● Transparency ● Robust Privacy Protocols	● Autonomous vehicles,	● Transport	● Lack of privacy ● Lack of consent
Mylrea and Robinson (2023)	USA	● Trust ● Security	● Evaluating privacy protection mechanisms in AI systems	● Health ● Finance	● Lack of privacy ● Lack of consent
Nasir et al. (2023)	Pakistan	● Transparency ● Human oversight	● Processing and analysing large amounts of patient medical data	● Health	● Lack of transparency
Genovesi and Mönig (2022)	Germany	● Respect for human autonomy. ● Fairness	● AI certification	● Environment	● Lack of Regulatory Compliance
Solanki et al. (2023)	Australia	● Privacy ● Security ● Transparency	● Predicting chronic diseases like type 2 diabetes	● Health	● Lack of privacy
Khan et al. (2021)	India	● Transparency, ● Privacy ● Accountability ● Fairness	● Development of AI products	● Information Technology	● Lack of Privacy ● Lack of security, ● Lack of transparency, ● Lack of accountability ● Lack of consent
Brendel et al. (2021)	Germany	● Transparency	● Diagnostics ● Credit scoring. ● Fraud detection	● Finance ● Transport	● Lack of privacy
Milossi et al. (2021)	Greece	● Privacy ● Data rights ● Responsible use ● Autonomy ● Data protection ● Fairness,	● Finance ● Health ● Business Operations	● Finance ● Health	● Lack of privacy ● Lack of consent ● Lack of Regulatory Compliance
Eitel-Porter (2021)	Europe	● Good governance ● Responsible use	● Compliance and governance risks	● Business Operations	● Lack of Regulatory Compliance
Peters et al. (2020)	USA	● Responsible design and use	● Digital Mental Health.	● Health	● Lack of transparency

3.1 Context

Asia and Oceania both had a 14% contribution of papers each. This geographical analysis shows how Europe with 41% and America with 31% of the publications are key players in the use and development of AI technologies. Africa and did not contribute any papers to the study (see Figure 2). The higher publication rates of AI frameworks in Europe and North America can be attributed to various factors such as investments in Research and Development. Europe and North America have made significant investments in artificial intelligence research, including funding dedicated to studying the ethical implications of AI systems (Bughin et al. 2018; Cath et al. 2018).

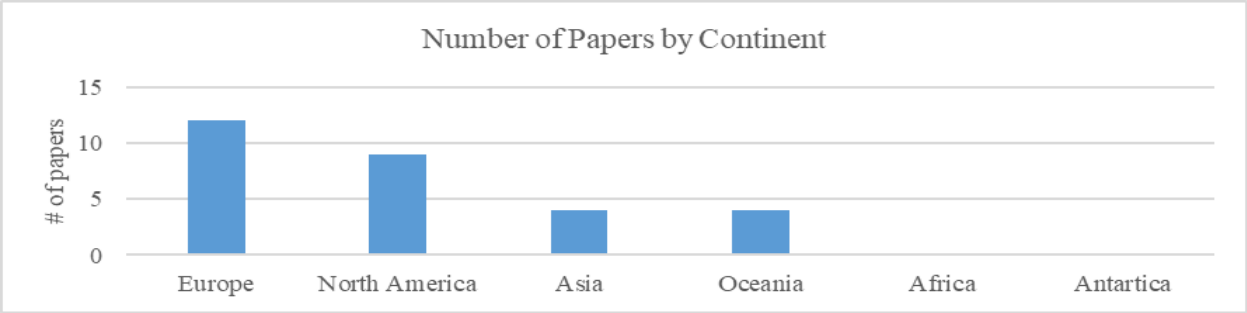

Figure 2. Distribution of papers by continent

Europe contributed 41 % of the papers. America was the second leading continent contributing 31% of the papers. These regions are home to prominent research centres, universities, and technology companies actively engaged in AI ethics research (Jobin et al. 2019). Europe, particularly with regulations like the General Data Protection Regulation (GDPR) and the proposed Artificial Intelligence Act, has led the way in developing frameworks for AI ethics (European Commission 2021). In North America, government-led initiatives and guidelines have also driven focus on ethical AI development. Researchers in Europe and North America have better access to funding, research infrastructure, and collaborative networks for ethical AI research due to the concentration of technology companies and research institutions in these regions (Bostrom and Yudkowsky 2014). As alluded to by Dwivedi et al. (2019), the lack of contributions from Africa and Antarctica can be explained by Limited AI Research, and Development and less established AI research ecosystems compared to Europe and North America due to limited resources, funding, and technical expertise. Developing regions like parts of Africa may prioritise addressing immediate socioeconomic challenges over investing in AI ethics research. Antarctica focused on scientific research and exploration, may have a lower emphasis on AI ethics. The geographical isolation and logistical difficulties of conducting research in Antarctica, as well as infrastructural and connectivity challenges in certain parts of Africa, may hinder the ability to participate in global research collaborations and publish in mainstream academic journals (Molinaro et al. 2016).

3.2 Key Features of AI Frameworks

The topic of AI frameworks is increasingly getting attention over the years with 79% of publications being done in 2023 and 2024 (see Figure 3) and less than 21 % of the papers were published between 2020 and 2022. The selected studies originated from institutions across four continents with a total of 14 countries. Transparency emerged as the most emphasised main feature, with 48.3% of the documents addressing this aspect. As alluded to by Singhal et al. (2024) transparency involves understanding the decision-making processes of AI systems. A total of 37.9% of the documents delved into privacy, making it the most addressed issue.

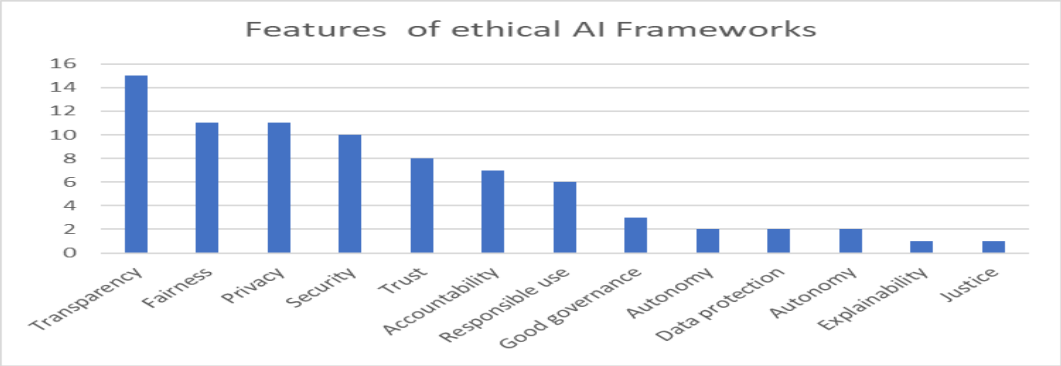

Figure 3. Features of ethical AI Frameworks.

According to Oluwabukunmi Latifat Olorunfemi et al. (2024), data privacy pertains to individuals' control over their personal information. Milossi et al. (2021) raised issues of privacy and data security regarding the collection and use of patient data for AI-driven applications since these systems must be trained with a large amount of personal data.

Data privacy concerns stem from factors such as data usage, ownership, and custodianship. Security (27.6%) and Trust (31%) Also demonstrated comparatively high frequencies compared to other features. According to Nasir et al. (2024) lack of transparency of AI algorithms, especially deep learning models, presents challenges in explaining their reasoning, potentially undermining trust in these systems.

3.3 Tools used to evaluate the ethical soundness of AI frameworks

Four studies namely Singhal et al. (2024); Nasir et al. (2023); Brendel et al. (2021); Prem (2023) made use of tools to evaluate the ethical soundness of AI frameworks. Specifically, Singhal et al. (2024) used Bias detection and deep learning. There is a lack of tools that can be used to evaluate the ethical soundness of AI frameworks with only 86.2% (Oluwabukunmi Latifat Olorunfemi et al. (2024); Williamson and Prybutok (2024); Lu et al. (2024) Kumar et al. (2024); Korobenko et al. (2024); Oluwaseun Augustine Lottu et al. (2024); Cumming et al. (2024); Pant et al. (2024); Saheb (2024); Obreja et al. (2024); Rai (2024); Coiduras-Sanagustín et al. (2024); Klarin et al. (2024); Chen et al. (2023); Tahri Sqalli et al. (2023); (Radanliev et al. 2024); Schmid and Wiesche (2023); Mylrea and Robinson (2023); Genovesi and Mönig (2022); Solanki et al. (2023); Khan et al. (2021); Eitel-Porter (2021); Peters et al. (2020); Li (2023) and Milossi et al. (2021)) of the papers not utilising any tools.

3.4 Applications areas and Domains of AI frameworks

Fourteen application areas were identified. Healthcare is the most prominent application area with 55% of the research papers with frameworks addressing various aspects like disease surveillance, patient care management, medical training, and diagnostics. Finance and Banking had a total of 21% of the papers that discussed AI frameworks that are used for credit scoring, fraud detection, and personalised financial services. Education and Transport had 21% whereas Information Technology had 10%. Governance, Legal and Justice, Public Administration, Security and Defence, social media, Environment, Agriculture, and Sociology all had 3% as displayed in Figure 4.

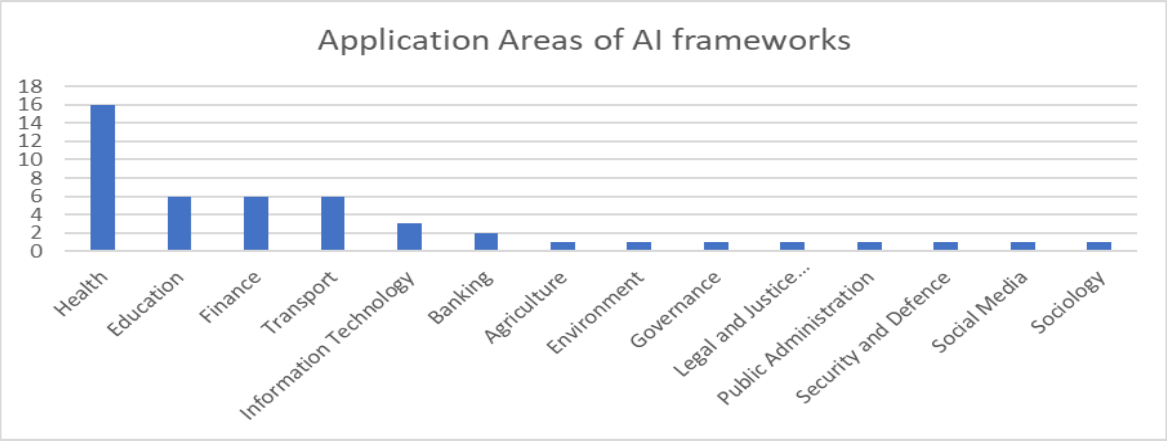

Figure 4. Application areas of AI frameworks

3.5 Challenges surrounding the privacy and security of AI frameworks

Challenges surrounding the privacy and security of AI frameworks highlight seven key concerns. A major issue is the lack of transparency (41.4%) in how AI systems operate, which leads to their potential data misuse and impacts on user privacy and consent. The studies reviewed highlight accountability (10.3%), insufficient data integrity (6.9%), and failure to comply with regulations (34.5%) as key challenges. There are also direct risks identified, such as security vulnerabilities (24.1%) and breaches of sensitive user information. Based on these challenges there is a need for more rigorous governance, oversight, and user control in the development and deployment of AI technologies. Empowering user choice, ensuring data privacy, and maintaining transparency are crucial to building trust and mitigating misuse of AI frameworks. Addressing these ethical and technical concerns is essential as AI becomes more widespread across different industries and application areas.

3.6 Dimension

Fifteen dimensions were identified in this study and were mentioned 50 times in the studied papers. Health was more dominant with 16 hits, translating to 32%, followed by finance with 8 hits translating to 16%. Education ranked third with 7 hits translating to 14%. Transport had 6 hits translating to 12%, information technology had 2 hits translating

to 4 % while agriculture, business operations, environment, governance, legal systems, organisational systems, public administration, security and defence, social media and sociology had 1 hit each, contributing to 2% each and 22% collectively.

4. Discussion

This section will provide a comprehensive discussion of the research findings and identify research gaps for future research.

4.1 Key features of AI frameworks

Solanki et al. (2020); Prem (2021) emphasises, that ensuring transparency and explainability is paramount in building ethical AI systems, and privacy concerns associated with AI technologies. Genovesi and Mönig (2022) allude that at least informed consent from users is needed about the collection and usage of their data by AI systems. The researchers propose consent models, which can be dynamic, enabling users at any point in time to review and update their consent preferences. Klarin et al. (2024) recommend granular choices, where users can select data, they are willing to share and for what purposes. Moreover, Nasir et al. (2024) provoked the usage of contextual consent models specific to the usage context, accompanied by appropriate choices for consent with corresponding user education programs to let them know the implications of their decisions. It is the findings of these studies that create the extent to which AI developers need to adopt empowerment and meaningful consent models that give users control over personal data.

4.2 Application Areas of AI Frameworks

Healthcare is the most prominent application area with 45% of the research papers with frameworks addressing various aspects like disease surveillance, patient care management, medical training, and diagnostics. Based on the research AI Ethics in healthcare focuses on patient privacy, data security, and transparency. Finance and Banking had a total of 21% of the papers discussed AI frameworks that are used for credit scoring, fraud detection, and personalized financial services. Key ethical concerns include fairness, data privacy, and security as alluded to by (Oluwabukunmi Latifat Olorunfemi et al. 2024). AI frameworks in the education sector were applied in personalised learning, educational assessment, and student support systems. Based on work done by Cumming et al. (2024), it highlights that ethical considerations revolve around fairness, transparency, security, and consent. In the transport sector, AI frameworks were used in autonomous vehicles, traffic management, and logistics optimization as seen in 8% of the papers analysed. Li (2023) highlights that safety, security, and bias in decision-making are the main ethical challenges in the transport sector. Saheb (2024) applied ethical frameworks in legal research, risk assessment, and sentencing recommendations. Ethical considerations include fairness, transparency, and potential for bias as cited by (Saheb 2024). AI frameworks were also used in various other fields, including social media, research and development, agriculture, manufacturing, and entertainment. Ethical considerations vary depending on the specific application. AI frameworks are being used in all domains, but the use of AI raises important ethical concerns that need to be addressed by a holistic framework.

4.3 Limitations of existing AI frameworks

The existing AI frameworks reviewed by the researchers have focused on key features such as privacy, transparency, fairness, trust, security, accountability, responsible use, good governance, autonomy, data protection, and explainability. However, the review has identified several limitations in these frameworks. This study noted a lack of a holistic approach to balancing ethical considerations, privacy, consent, and responsible use. Milossi et al. (2021), eluded that the frameworks exhibited an insufficient focus on user privacy, data consent, and control mechanisms. Furthermore, Lu et al. (2024) observed a limited use of established theoretical frameworks or models to guide the development of these ethical AI frameworks. The review also highlighted the inadequate use of evaluation tools to assess the ethical soundness and practical implementation of the proposed frameworks, restricting the ability to validate their effectiveness. To address these limitations, the researchers proposed a comprehensive and holistic approach to developing a new ethical AI framework. The proposed framework should incorporate a stronger emphasis on user privacy, data consent, and control mechanisms.

This study also highlighted the importance of expanding the tools used to evaluate the ethical soundness of AI frameworks. A few researchers used tools, including Bias Detection Algorithms Singhal et al. (2024), Deep Learning techniques Nasir et al. (2023), and Natural Language Processing Brendel et al. (2021); Prem (2023), to assess the ethical soundness and practical implementation of the proposed framework.

4.4 Challenges surrounding the privacy and security of AI frameworks?

The primary concern in AI frameworks, privacy, security, consent, and responsible use, arises from the lack of data privacy and security safeguards. Many of the studies (80%) highlight that AI applications require big datasets for their operations, making them targets of cyber-attacks that can result in breaches of sensitive personal data. As discussed by Singhal et al. (2024) the challenge lies in security vulnerabilities. Non-transparent and non-explainable AI systems add to the complexity and opaqueness of many AI models, wherein it is hard to understand how decisions are made, this raises several questions regarding accountability and obtaining informed consent as alluded to by (Cumming et al. 2024). The research results point to concern about bias and fairness, AI systems may further intensify certain biases and inequalities in society, thereby decreasing how technologies are responsibly and fairly used as discussed by (Oluwabukun mi Latifat Olorunfemi et al. 2024). Critical challenges highlighted in this research propose strengthening data privacy and security, transparency, and explainability, mitigating bias to ensure that responsible development and use of AI frameworks respect individual rights and foster ethical outcomes.

Most AI models are incredibly complex, making it impossible to understand how the decisions have been made, this reduces accountability within a framework and user trust (Cumming et al. 2024). AI frameworks should be designed to offer greater overall transparency; for example, using explainable AI techniques that provide insight into the reasoning behind predictions and decisions. Artificial intelligence frameworks must integrate bias detection and mitigation strategies, to ensure the fairness and equitability of results the researchers identified the need to include regulatory and legal frameworks to guide the development and deployment of AI. The governance structure is necessary to make sure AI is being used responsibly, in sensitive domains like healthcare and finance. This will address ethical concerns. Researchers identified the need for users to be educated on how their data is being collected and used and be able to exercise control. Addressing these concerns in the new proposed AI framework can ensure responsible development and use, respecting individual rights whilst fostering ethical outcomes.

5. Proposed AI framework

The ethical framework proposed in this study includes AI privacy, security (Figure 5), transparency, accountability, fairness, consent, compliance, and governance. The AI framework will lay out practical ways of ensuring that ethical concerns are consistently put into practice within real-world settings. This framework would be dynamic, meaning it changes as technology does but remains constant and faithful to social ethics. It has strong privacy and security settings on data, especially for sensitive industries like healthcare. This makes compliance with regulations easier to adopt and use. In winning public trust transparency, accountability and fairness play a big role. Apart from that, it would establish processes and tools easily executable for stakeholders, insert sustainability considerations into driving green and socially responsible practices, and enhance security with proper protection measures against diverse risks to enable the safe deployment of AI systems. This comprehensive Ethical AI Framework aims to address the key ethical concerns and regulatory requirements identified in the studies by (Ndlovu et al. 2022). By incorporating these essential elements, the framework strives to balance privacy, consent, security, and responsible use.

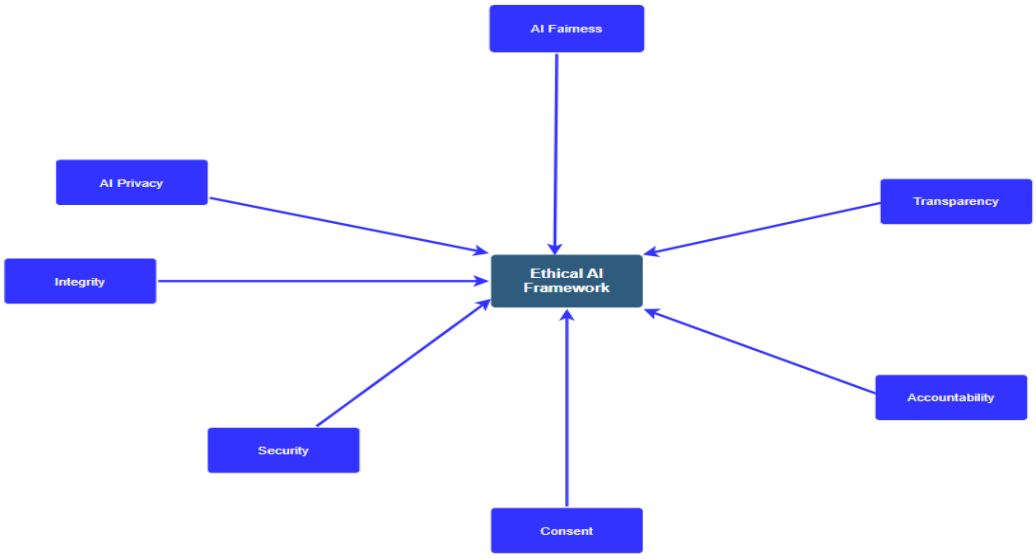

Figure 5. Proposed Holistic AI Framework

6. Conclusion Recommendations and Future Works

The development and deployment of artificial intelligence (AI) systems raise a lot of ethical concerns. To ensure that AI benefits society, it is crucial to build trust by fostering ethical considerations throughout the AI lifecycle. The researcher analysed existing studies on ethical AI frameworks, highlighting the importance of Privacy, consent, transparency, responsible use, accountability, and fairness. The research identified challenges in developing and deploying AI systems. Recommendations include implementing robust data privacy and security measures Maguraushe et al. (2024), enhancing transparency and explainability in AI models, mitigating bias and fairness issues, establishing governance and accountability mechanisms, and empowering users with choices and transparency about their data usage. Establishing robust data privacy and security measures is crucial for AI systems, including strong encryption, fine-grained access controls, and data minimization techniques. Enhancing transparency and explainability in AI models can increase accountability and user trust. AI frameworks should incorporate bias detection and mitigation strategies to mitigate bias and fairness issues. Robust governance and accountability mechanisms, such as regulatory frameworks and oversight boards, are necessary for ethical and responsible AI use. Empowering users with choices and transparency about their data usage can enhance user autonomy and alleviate concerns about misuse. Building on these studies, the researchers proposed a new holistic AI framework. The proposed framework provides a springboard for further empirical research and development in this critical area. This proposed framework can be integrated with other frameworks such as the one proposed by Dube et al. (2023) As AI continues to evolve, so must our efforts to ensure its ethical and socially responsible application.

References

Bostrom, N., & Yudkowsky, E, The ethics of artificial intelligence. In F. Ramsey (Ed.), The Cambridge Handbook of Artificial Intelligence. Cambridge University Press, 2014.

Brendel, A. B., Mirbabaie, M., Lembcke, T. B., & Hofeditz, L, Ethical management of artificial intelligence, In *Sustainability,* Vol. 13, Issue. 4, pp. 1–18, 2021.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., ... & Trench, M, Notes from the AI frontier: Insights from hundreds of use cases. McKinsey Global Institute, 2018.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. Science and Engineering Ethics, Vol. 24, Issue. 2, pp 505-528, 2018.

Chen, F., Zhou, J., Holzinger, A., Fleischmann, K. R., & Stumpf, S, Artificial Intelligence Ethics and Trust, From Concerns to Practice, Institute of Electrical and Electronics Engineers Inc, Vol. 38, Issue. 6, pp. 5–8, 2023.

Coiduras-Sanagustín, A., Manchado-Pérez, E., & García-Hernández, C, Understanding perspectives for product design on personal data privacy in the internet of things (IoT): A systematic literature review (SLR). *Heliyon*, Vol. *10, Issue.* 9, 2024.

Cumming, D., Saurabh, K., Rani, N., & Upadhyay, P, Towards AI ethics-led sustainability frameworks and toolkits, Review and research agenda, *Journal of Sustainable Finance and Accounting*, Vol. *1, 2024*.

Dube, S., Dube, S., Ndlovu, B. M., Maguraushe, K., Malungana, L., Kiwa, F. J., & Muduva, M, *Students' Perceptions of* ChatGPT in Education: A Rapid Systematic Literature Review, pp 258–279, 2024. In: Arai, K. (eds) Intelligent Computing. SAI 2024. Lecture Notes in Networks and Systems, vol 1019. Springer, Cham. https://doi.org/10.1007/978-3-031-62273-1_18

Dube, S., Mutunhu, B., & Dube, S. P., Lecturers' Experiences in Teaching STEM Courses Online During COVID-19: Case of a Zimbabwean University. *The IAFOR Conference on Educational Research & Innovation: 2023 Official Conference Proceedings*, 53–65.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice, and policy. International Journal of Information Management, Vol. 47, pp 123-129, 2019.

Eitel-Porter, R. Beyond the promise: implementing ethical AI. *AI and Ethics*, Vol. *1, Issue* 1, pp 73–80, 2021.

European Commission. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. COM (2021) 206 final.

Florida, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E, AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Concerns and Recommendations, *Minds and Machines*, Vol. *28, Issue* 4, pp 689–707, 2018.

Genovesi, S., & Mönig, J. M, Acknowledging Sustainability in the Framework of Ethical Certification for AI. *Sustainability (Switzerland)*, Vol. *14, Issue.* 7, 2022.

Jobin et al. A., Ienca, M., & Vayena, E., The global landscape of AI ethics guidelines. Nature Machine Intelligence, Vol.1, Issue. 9, pp 389-399, 2019.

Khan, A. A., Badshah, S., Liang, P., Khan, B., Waseem, M., Niazi, M., & Akbar, M. A., *Ethics of AI: A Systematic Literature Review of Concerns and Challenges, 2021*.

Klarin, A., Ali Abadi, H., & Sharmelly, R. Professionalism in artificial intelligence, The link between technology and ethics, *Systems Research and Behavioral Science, 2024*.

Korobenko, D., Nikiforova, A., & Rajesh, S, *towards a Privacy and Security-Aware Framework for Ethical AI, Guiding the Development and Assessment of AI Systems*, 2024.

Kumar, A. N. P., Bogner, J., Funke, M., & Lago, P., *Balancing Progress and Responsibility: A Synthesis of Sustainability Trade-Offs of AI-Based Systems, 2024*.

Li, N. Ethical Considerations in Artificial Intelligence, A Comprehensive Discussion from the Perspective of Computer Vision, *SHS Web of Conferences*, Vol. *179*, 2023.

Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A, Responsible AI Pattern Catalogue: A Collection of Best Practices for AI Governance and Engineering. *ACM Computing Surveys*, Vol. *5, Issue. 7*, 2024.

Maguraushe, K., da Veiga, A., & Martins, N, A personal information privacy perceptions model for university students. Information Security Journal: A Global Perspective, Vol. 33, Issue 4, pp 394–424, 2024. https://doi.org/10.1080/19393555.2024.2329554

Milossi, M., Alexandropoulou-Egyptiadou, E., & Psannis, K. E, AI Ethics: Algorithmic Determinism or Self-Determination? The GPDR Approach. *IEEE Access*, Vol. *9*, 2021.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G, Preferred reporting items for systematic reviews and meta-analyses, the PRISMA statement, Vol. *339*, Issue. 7716, pp 332–336, 2009.

Mutunhu, B., Dube, S., Dube, S. P., & Mpofu, S. *A Framework for Transitioning to Virtual Classes During Life-Threatening Pandemics Like COVID-19*. In European Conference on e-Learning (Vol. 21, No. 1, pp. 279-287), 2022.DOI: https://doi.org/10.34190/ecel.21.1.900

Mutunhu, B., Dube, S., Ncube, N., & Sibanda, S, *Cyber Security Awareness and Education Framework for Zimbabwe Universities: A Case of National University of Science and Technology, 2022*.

Mylrea, M., & Robinson, N. Artificial Intelligence (AI) Trust Framework and Maturity Model: Applying an Entropy Lens to Improve Security, Privacy, and Ethical AI, Vol. *25, Issue.* 10, 2023.

Nasir, S., Khan, A., Bai, S., & Khan, R. A*., Ethical Framework for Harnessing the Power of AI in Healthcare and Beyond, 2024*.

Obreja, D. M., Rughiniș, R., & Rosner, D., Mapping the conceptual structure of innovation in artificial intelligence research: A bibliometric analysis and systematic literature review. *Journal of Innovation and Knowledge*, Vol. *9, Issue.* 1, 2024.

Oluwabukunmi Latifat Olorunfemi, Olukunle Oladipupo Amoo, Akoh Atadoga, Oluwatoyin Ajoke Fayayola, Temitayo Oluwaseun Abrahams, & Philip Olaseni Shoetan. TOWARDS A CONCEPTUAL FRAMEWORK FOR ETHICAL AI DEVELOPMENT IN IT SYSTEMS. *Computer Science & IT Research Journal*, Vol. *5, Issue.* 3, pp 616–627, 2024.

Oluwaseun Augustine Lottu, Boma Sonimiteim Jacks, Olakunle Abayomi Ajala, & Enyinaya Stefano Okafor. Towards a conceptual framework for ethical AI development in IT systems, *World Journal of Advanced Research and Reviews*, Vol. *21, Issue.* 3, pp 408–415, 2024.

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., Mcdonald, S., … Mckenzie, J. E, *Supplementary Material to: PRISMA 2020 explanation and elaboration: updated guidance and examples for reporting systematic reviews, 2020*.

Pant, A., Hoda, R., Tantithamthavorn, C., & Turhan, B., Ethics in AI through the practitioner's view: a grounded theory literature review. *Empirical Software Engineering*, Vol. *29, Issue* 3, 2024.

Peters, D., Vol.d, K., Robinson, D., & Calvo, R. A. Responsible AI—Two Frameworks for Ethical Design Practice, *IEEE Transactions on Technology and Society*, Vol. *1, Issue* 1, pp 34–47, 2023.

Prem, E. From ethical AI frameworks to tools: a review of approaches. *AI and Ethics*, *3*(3), 699–716, 2020, 2023,

Radanliev, P., Santos, O., Brandon-Jones, A., & Joinson, A., Ethics and responsible AI deployment. In *Frontiers in Artificial Intelligence* Vol. 7, 2024.

Rai, P. *Ethics in AI: A Deep Dive into Privacy Concerns,* 2023.

Saheb, T, Mapping Ethical Artificial Intelligence Policy Landscape: A Mixed Method Analysis. *Science and Engineering Ethics*, Vol. *30, Issue* 2, 2024.

Schmid, A., & Wiesche, M, The Importance of an Ethical Framework for Trust Calibration in AI. *IEEE Intelligent Systems*, Vol. *38, Issue.* 6, pp 27–34, 2023.

Singhal, A., Neveditsin, N., Tanveer, H., & Mago, V. (2024), Toward Fairness, Accountability, Transparency, and Ethics in AI for social media and Health Care: Scoping Review. In *JMIR Medical Informatics* (Vol. 12), 2024.

Solanki, P., Grundy, J., & Hussain, W, Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. *AI and Ethics*, Vol. *3, Issue* 1, pp 223–240, 2023.

Tahri Sqalli, M., Aslonov, B., Gafurov, M., & Nurmatov, S, Humanising AI in medical training: ethical framework for responsible design. *Frontiers in Artificial Intelligence*, Vol. *6*, 2023.

UK should be more positive about AI to avoid missing out on tech 'goldrush', The Independent, July 2024. https://www.independent.co.uk/business/uk-should-be-more-positive-about-ai-to-avoid-missing-out-on-tech-goldrush-b2489172.html

Williamson, S. M., & Prybutok, V, Balancing Privacy and Progress: A Review of Privacy Challenges, Systemic Oversight, and Patient Perceptions in AI-Driven Healthcare. In *Applied Sciences (Switzerland)*, Vol. 14, Issue 2, 2024.

Biographies

Belinda Ndlovu is a PhD candidate in Information Systems at UNISA. She is also a full-time lecturer at the National University of Science and Technology. She holds a Master of Science in Information Systems, a BSc in Computer Science, and a Post Grad Diploma in Education.

Sibusisiwe Dube is a Professor at the National University of Science and Technology in Zimbabwe. She holds a BSc Hons in Information Systems from the Midlands State University in Zimbabwe, an MSc in Computer Science from the National University of Science and Technology in Zimbabwe, and a PhD in Information Systems from the University of Cape Town in South Africa.

Tafadzwa Mukudu is a master's degree student in the Faculty of Applied Science Department of Informatics at the National University of Science and Technology in Zimbabwe. She holds a Bachelor of Science Honors degree and a Master of Science in Operations Research and Statistics from the National University of Science and Technology. Currently, she works as a Data Scientist.

Mqamlandaba Dube is currently studying MSc in Big Data Science at the National University of Science and Technology in Zimbabwe (NUST). He holds a BSc honours degree in informatics from (NUST) and Applied Data Science 1: Scientific Computing and Python (with honours) from WorldQaunt University. Diploma in Electrical Engineering, Diploma in project management.

Tawanda Kapuya is a Master of Science in Big Data student at the National University of Science and Technology. He holds a Bachelor of Science degree in Computer Science from Great Zimbabwe and works as a System Administrator at Turbo Mining Pvt Ltd. He develops scripts, software, and various software configurations spanning from custom-made to off-the-shelf. He has vast networking experience and doubles as a Network Administrator in his current role.

Mind Kutyauripo is currently a master's in Big Data student at the National University of Science and Technology.

Walter Kubiku holds a Bachelor of Science Honours in Computer Science from the National University of Science and Technology. He is pursuing his Master of Science in Big Data at the same institution. Walter is engaged at TV Sales and Home as a Helpdesk Support Analyst, tasked with analysing system operations and performance and user support. His skills include ERP: Open Bravo, Website: WooCommerce, Network: Sophos, and Windows Server: Active Directory. He has experience with VOIP systems.