



Exploring Ethical Implications of Artificial Intelligence (AI) Algorithms in Decision-Making: A Human-Centric Approach

Christopher Andere Agwenyi, Anselemo Peters Ikoha

Department of Information Technology, School of Computing and Informatics, Kibabii University, P.O Box1699-50200, Bungoma, Kenya

DOI: <https://doi.org/10.68050/JAMS.2026.654>

Received: 06 September 2026; Accepted: 12 September 2026; Published: 01 October 2026

ABSTRACT

Artificial Intelligence (AI) algorithms are increasingly being integrated into decision-making processes across various sectors, raising critical ethical concerns. This paper explores the ethical implications of AI algorithms in decision-making and advocates for a human-centric approach to AI governance. We analyze key ethical challenges, discuss frameworks for AI ethics, ethical dilemmas and propose strategies to ensure human values and rights are preserved in AI-driven decision systems.

Keywords: Artificial Intelligence, Ethics, Decision-Making, AI Algorithms, Human-Centric Approach

INTRODUCTION

In recent years, artificial intelligence (AI) and machine learning (ML) systems have transitioned from theoretical decision-support models to active clinical tools integrated directly into healthcare workflows. Across complex medical domains ranging from oncological image analysis to neuro-rehabilitation AI promises to personalize therapy, optimize resource allocation, and enhance diagnostic accuracy. However, as algorithmic systems assume higher levels of agency in patient care, critical concerns surrounding ethics, algorithmic bias, model opacity, and clinical responsibility have intensified. High-stakes application areas, such as stroke rehabilitation, illustrate these tensions: while AI-driven predictive tools can tailor long-term motor recovery plans, their underlying decision-making mechanisms often function as "black boxes," leaving clinicians and patients uncertain about how individual recovery trajectories or risk assessments are generated.

Despite a growing corpus of scholarly literature advocating for "responsible," "ethical," or "human-centric" AI, the translation of high-level normative principles into operational clinical workflows remains fraught with ambiguity. Existing scholarship frequently conflates closely related ethical constructs such as algorithmic fairness, transparency, explainability, accountability, human oversight, and human-AI collaboration treating them as interchangeable ideals rather than distinct operational dimensions. Consequently, clinicians, healthcare administrators, and AI developers lack a unified, multi-layered governance structure capable of balancing technical capabilities with clinical autonomy, regulatory oversight, and patient-centered ethics.

To address these challenges, this manuscript presents a **conceptual framework-development literature review**. Grounded in a thematic analysis of interdisciplinary scholarship spanning computer science, bioethics, healthcare management, and clinical rehabilitation, this paper bridges the gap between high-level ethical abstractions and practical clinical deployment.

The primary objective of this study is to synthesize existing interdisciplinary research into an integrated, human-centric AI governance framework for clinical decision-support systems, using stroke rehabilitation as a primary illustrative domain. Specifically, this paper addresses the following three Research Questions (RQs):

- i. **RQ1:** How are key ethical concepts specifically algorithmic fairness, transparency, explainability, accountability, human oversight, and human-AI collaboration defined, differentiated, and operationalized within healthcare AI literature?



- ii. **RQ2:** How do clinical case studies, particularly in AI-assisted stroke rehabilitation, illuminate the operational friction and ethical risks that arise when algorithmic predictions intersect with real-world clinical decision-making?
- iii. **RQ3:** What architectural model can integrate regulatory, technical, and human-in-the-loop mechanisms into a cohesive, human-centric AI governance framework for high-stakes healthcare environments?

LITERATURE REVIEW

Previous studies have highlighted various ethical challenges associated with AI algorithms, including algorithmic bias, lack of transparency, and the potential for unintended consequences. Scholars advocate for ethical AI design principles, such as fairness, accountability, transparency, and human oversight. Additionally, frameworks like the IEEE Ethically Aligned Design provide guidelines for integrating ethical considerations into AI development.

As the performance of AI systems has rapidly improved to match or exceed that of human experts' people have considered substituting human decision making with predictions of these systems in a variety of applications. However, deploying fully autonomous AI systems remains disruptive, dangerous, and unethical in high-stake contexts. Instead, the need of involving a human to interact with AI systems has received increasing attention. Amershi et al., (2019) propose design guidelines for human-AI interaction through a user study with 49 design practitioners. According to Schneiderman, (2020) presents ethical principles and practical steps for human centered AI systems. Building upon these guidelines, we focus on exploring human-AI collaboration in the high-stakes context of clinical decision making.

Clinical decision support systems have been considered as promising ways that can provide medical practitioners computational information on the status of a patient to improve their decision making on various disciplines (e.g. cancer diagnosis, detection of diabetic retinopathy, or assessment of rehabilitation therapy. However, even if such systems have the potential to improve the quality and efficiency of health care the adoption of these systems in practice remains a challenge due to the lack of user-centered designs and the opaqueness of machine learning algorithms.

For better deployment, recent research efforts in clinical decision support systems have demonstrated the value of involving the end user in the process of design and evaluation. Yang et al., (2020) conducted a field evaluation on the design of a decision support tool for cardiologists with synthetic data, and found that clinicians are more likely to embrace a tool that augments their decision making in natural and intuitive ways. interviewed pathologists about their desires about an AI assistant for prostate cancer diagnosis, and discussed that one major need is to make an AI assistant transparent by informing its overall capability and limitation on a task.

AI is turning up everywhere, but when people try to use it to as a tool help make better decisions it often stumbles. . .users reject it, rely too much on it, and so forth. Not only is this a problem for AI as-tool applications, it is increasingly clear that AI without human oversight is prone to bad mistakes, typically because the AI has such a narrow view of the world and can't tell when it is violating norms or when the context has changed. As a consequence, AI-automation is getting serious pushback from citizens and lawmakers.

Looking at previous works, we have seen how people perceived AI making ethical decisions. Using a simulation for decision making, experiments conducted that investigated how people's perceptions for human experts versus AI differed on 1) trust they place in the expert, 2) responsibility they ascribe to the expert, and 3) reliance they show on the expert. We deduced that variables across different framing of the ethical dilemmas and for different levels of autonomy of the expert. We find that people show a higher capacity trust, overall trust, and reliance on AI experts, but have somewhat higher moral trust and higher responsibility ascription for human experts.



THEMATIC LITERATURE REVIEW

Rather than evaluating AI literature on a study-by-study basis, this section organizes scholarly discourse into six interconnected thematic pillars: **Algorithmic Fairness and Bias**, **Transparency**, **Explainability (XAI)**, **Accountability**, **Human Oversight**, and **Human-AI Collaboration**. A frequent shortcoming in existing AI governance discourse is the conflation of these terms, treating them as interchangeable normative goals. To establish a rigorous analytical foundation for framework development, Table 1 delineates these core concepts, establishing their precise definitions, clinical scopes, and potential failure modes when unaddressed.

Table 1: Conceptual Delineation Matrix of Core AI Governance Pillars

Concept	Core Definition	Clinical / Operational Scope	Primary Failure Mode / Risk
Algorithmic Fairness & Bias	Freedom from systemic demographic, socioeconomic, or procedural prejudice in model predictions.	Ensuring equitable performance across diverse patient subgroups (e.g., age, sex, ethnicity, disability baseline).	Disparate accuracy; exacerbation of health disparities in marginalized populations.
Transparency	System-level accessibility and visibility into data origins, architectural design, and operational boundaries.	Open documentation regarding training data provenance, model architecture, hyperparameter choices, and validation limits.	"Black-box" opacity; inability of health systems to conduct pre-deployment risk audits.
Explainability (XAI)	The capacity to interpret and present the rationale behind a specific algorithmic output in human-understandable terms.	Generating local interpretations (e.g., feature importance maps, counterfactuals) for specific patient risk scores or motor trajectories.	Automation bias or complete rejection of AI recommendations by mistrustful clinicians.
Accountability	Clear assignment of legal, ethical, and organizational responsibility for AI-driven or AI-assisted clinical outcomes.	Defining liability boundaries among software vendors, hospital administrations, and attending clinicians.	Diffusion of responsibility ("liability vacuum") when an AI-driven clinical recommendation leads to harm.
Human Oversight	Supervisory mechanisms and protocols that ensure human actors retain authority to override or constrain AI actions.	Protocol-driven approval gates ("human-in-the-loop") prior to executing AI-recommended therapy or treatment modifications.	Unchecked algorithmic agency; failure to intervene during edge-case or out-of-distribution events.
Human-AI Collaboration	Synergistic decision-making paradigms where human expertise and algorithmic analytics mutually reinforce each other.	Dynamic co-reasoning models combining clinician intuition/qualitative observations with AI predictive trajectory analytics.	Cognitively overwhelming interfaces; misalignment with established clinical workflows.

Decision-making algorithms are being used in important decisions, such as who should be enrolled in health care programs and can be hired. Even though these systems are currently deployed in high-stake scenarios, many of them cannot explain their decisions. This limitation has prompted explainable Artificial Intelligence initiative, which aims to make algorithms explainable to comply with the legal regulations, promote trust and maintain accountability.



We conclude that for AI for ethical decision making to become a reality, these differences in capabilities need to be accounted for in the design of the AI and decision-making process by exploring ethical implications of artificial intelligence (AI) algorithms in decision-making: a human-centric approach.

METHODOLOGY

This study was anchored on the premise of interpretivism research paradigm approach. Interpretivism, which is sometimes called Anti-Positivism is that knowledge in the human and social sciences cannot conform to the model of natural science because there are features of human experience that cannot objectively be “known”. This might include emotions; understandings; values; feelings; subjectivities; socio-cultural factors; historical influence; and other meaningful aspects of human being. Interpretivism tends to emphasize the subjective over the objective.

This study adopts a **systematic thematic synthesis methodology** tailored for conceptual framework development (adapted from Braun & Clarke, 2006; Whittemore & Knafl, 2005). Rather than presenting a descriptive summary of individual papers, this review systematically extracts, categorizes, and synthesizes multi-disciplinary literature to establish conceptual clarity and construct an actionable governance architecture.

We analyze existing literature, case studies, and policy documents related to AI ethics. Key themes and ethical considerations are identified and synthesized to inform our human-centric approach.

Qualitative methodology focuses on understanding human behavior within its natural context through non-numerical data collection methods such as interviews, focus groups, or observations. One major strength of qualitative methodology is its ability to provide rich and in-depth insights into complex phenomena that cannot be adequately captured by quantitative measures alone. It allows researchers to uncover nuances and understand participants' perspectives more holistically.

On the other hand, qualitative methodology has some limitations as well. The findings may lack generalizability due to the small sample sizes typically used. Additionally, data analysis in qualitative research can be time-consuming and subjective, requiring careful interpretation and coding.

Data collection methods included interviews, observations and focus group discussions (qualitative) to delve deeper into participants' experiences and perceptions. Qualitative data provided insights into implementation challenges and user experiences. Qualitative research methods are essential in understanding the complexities of computing research, focusing on the "how" and "why" of human behavior, organizational practices, and social phenomena related to technology. These methods enable researchers to explore the context and meaning of different computing environments, practices, and experiences.

Data analysis methods employed included; thematic analysis method for identifying, analyzing, and reporting patterns (themes) within data. It minimally organizes and describes your data set in (rich) detail. However, frequently it goes further than this, and interprets various aspects of the research topic. In computing research, thematic analysis can be employed to analyze qualitative data from usability studies, user experience research, or the social implications of technology deployment. Discourse analysis method used to analyze written, spoken, or sign language use, or any significant semiotic event. It goes beyond the content to examine the social and psychological implications of language use within a context. In the field of computing, it can be particularly valuable for studying online communications, technology-mediated interactions, and the cultural narratives surrounding technology use.

In conclusion, choosing an appropriate methodology for your research proposal is crucial for ensuring valid results that address your research question effectively. Whether you opt for quantitative, qualitative, mixed-methods, or experimental approaches will depend on various factors such as your research question, available resources, target population characteristics, and disciplinary conventions. By understanding the strengths and limitations of each methodology, one can make an informed decision that aligns with your research goals.

To ensure methodological rigor, transparency, and reproducibility in synthesizing literature and constructing the proposed governance framework, this study adopts a **systematic thematic synthesis methodology** tailored for conceptual framework development (adapted from Braun & Clarke, 2006; Whitemore & Knafl, 2005). Rather than presenting a descriptive summary of individual papers, this review systematically extracts, categorizes, and synthesizes multi-disciplinary literature to establish conceptual clarity and construct an actionable governance architecture as shown in Figure 3.1

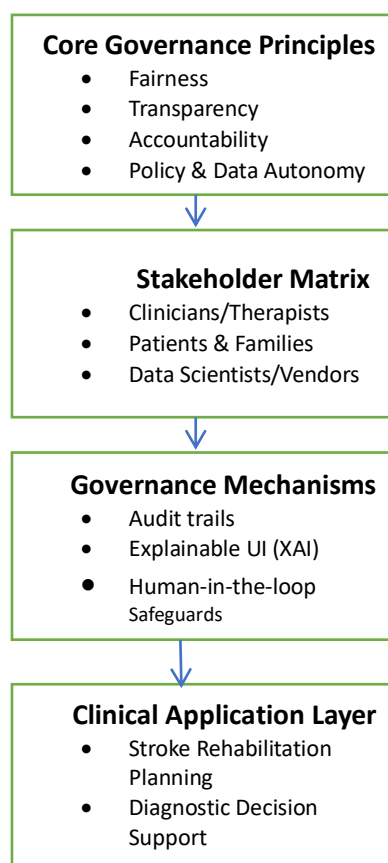


Figure 3.1 Human Centric AI Governance Framework

RESULTS

Ethical dilemmas related to privacy, freedom of speech, and censorship in IT research require careful consideration of the complex interplay between individual rights, societal interests, technological advancements, and cultural values. Researchers, policymakers, and technology practitioners must engage in ethical decision-making processes to navigate these dilemmas responsibly and uphold fundamental principles of human dignity, autonomy, and justice in the digital age; **Privacy:** Ethical Dilemma: Balancing the right to privacy with the need for data collection and analysis for research or commercial purposes. (Power, D. J., Heavin, C., & O'Connor, Y. ,2021). **Explanation:** As technology enables the collection and analysis of vast amounts of personal data, concerns arise regarding the ethical implications of surveillance, data mining, and profiling practices that may infringe on individuals' privacy rights. **Freedom of Speech:** Ethical Dilemma: Addressing hate speech and harmful content online while upholding the principles of free expression and open discourse. (Johnson, B. G. ,2018). **Explanation:** The proliferation of social media platforms and online forums has raised ethical concerns about the spread of misinformation, cyberbullying, and extremist ideologies, prompting debates about the appropriate limits of free speech in digital spaces. **Censorship:** Ethical Dilemma: Navigating the tension between protecting users from harmful content and preserving freedom of expression in online platforms. (Ullmann, S., & Tomalin, M. ,2020). **Explanation:** Content moderation practices employed by tech companies to combat hate speech, fake news, and other forms of harmful content raise ethical questions about censorship, algorithmic bias, and the power of platforms to shape public discourse.



Governments may engage in censorship for several reasons, which may vary depending on the political, social, and cultural context. Some of the main justifications for government censorship include: **National security:** Governments may censor information that they perceive as a threat to national security, such as classified military documents, sensitive intelligence operations, or terrorist propaganda (Stone, G. R. ,2009). Censorship in the name of national security aims to prevent the dissemination of information that could jeopardize the safety and stability of the country. **Public order and safety:** Censorship may be used to maintain public order and safety by restricting the spread of content that incites violence, promotes hate speech, or fuels social unrest. Governments may intervene to suppress inflammatory rhetoric or content that could lead to civil unrest or public disturbances. **Protecting public morals and values:** Governments may censor content that is deemed morally or culturally objectionable to uphold societal norms and values (Coetzee, J. M. ,2018). This can include censoring explicit or obscene material, pornography, or content that goes against religious or cultural beliefs. **Protecting children and vulnerable populations:** Censorship may be employed to shield children and vulnerable populations from exposure to harmful or inappropriate content (Robinson, K. H. ,2013). Governments may restrict access to certain types of media or online content that are considered unsuitable for minors, such as violence, pornography, or explicit language. **Preventing defamation and protecting reputation:** Censorship may be used to prevent the spread of false information, defamation, or libel that could harm individuals' or institutions' reputations (Land, M. K.,2019). Governments may intervene to suppress defamatory content or misinformation that could damage public trust or undermine confidence in public figures or institutions. **Regulating information flow:** Governments may engage in censorship to regulate the flow of information and control the narrative on certain issues. Censorship can be used to suppress dissenting voices, manipulate public opinion, or maintain political control by controlling access to information and shaping public discourse.

While these reasons may be invoked to justify government censorship, it is important to recognize that censorship can also be abused to suppress dissent, stifle free speech, and infringe on individual rights and freedoms. Therefore, the use of censorship by governments should be carefully scrutinized and balanced with respect for fundamental human rights, including the right to freedom of expression and access to information. **Data Privacy and Security:** Ethical Dilemma: Balancing the benefits of data-driven innovation with the risks of privacy breaches and unauthorized access to sensitive information. **Explanation:** The collection, storage, and sharing of personal data by corporations and governments raise ethical concerns about consent, transparency, and accountability in data practices, particularly in the context of emerging technologies like artificial intelligence and the Internet of Things. **Digital Divide:** Ethical Dilemma: Addressing disparities in access to technology and digital skills that exacerbate social inequalities and limit opportunities for marginalized communities. (Helsper, E. ,2021). **Explanation:** Unequal access to digital resources and opportunities perpetuates social and economic disparities, raising ethical questions about fairness, inclusivity, and social justice in the digital age.

In summary, ethical dilemmas related to privacy, freedom of speech, censorship, data privacy and security, and the digital divide highlight the complex interplay between individual rights, societal values, technological advancements, and regulatory frameworks in the realm of IT research and practice.

DISCUSSION

The discussion section delves into the importance of adopting a human-centric approach to AI governance. We propose strategies for enhancing transparency, accountability, and fairness in AI systems. Collaborative efforts involving policymakers, technologists, ethicists, and civil society are essential to ensure AI algorithms align with human values and rights.

Decision-making algorithms are being used in important decisions, such as who should be enrolled in health care programs and can be hired. Even though these systems are currently deployed in high-stake scenarios, many of them cannot explain their decisions. This limitation has prompted explainable Artificial Intelligence initiative, which aims to make algorithms explainable to comply with the legal regulations, promote trust and maintain accountability.



We conclude that for AI for ethical decision making to become a reality, these differences in capabilities need to be accounted for in the design of the AI and decision-making process by exploring ethical implications of artificial intelligence (AI) algorithms in decision-making: a human-centric approach.

The synthesis of literature and the formulation of the Human-Centric AI Governance Framework underscore a fundamental reality in health informatics: technical performance metrics (e.g., AUC-ROC, accuracy, F1-scores) are necessary but wholly insufficient for safe, equitable, and effective clinical AI integration. By moving beyond isolated discussions of algorithmic properties, this study addresses the structural gaps identified in existing literature through three major theoretical and practical contributions.

Theoretical Implications: De-siloing Ethical Pillars

A primary contribution of this work is the explicit conceptual delineation and structural integration of core AI governance pillars (Table 1). Prior scholarship has frequently treated constructs such as *explainability*, *fairness*, and *oversight* as distinct, independent research agendas. In practice, however, these pillars are deeply interdependent:

- i. **Explainability as an Enabler of Fairness:** Without local explainability (e.g., feature importance maps or counterfactual reasoning), clinicians cannot verify whether a predictive model is relying on clinically valid indicators or spuriously correlating demographic proxies with adverse outcomes.
- ii. **Oversight as an Operationalization of Accountability:** Human oversight cannot function meaningfully in a vacuum. Effective "human-in-the-loop" safeguards require clear legal and organizational frameworks that define clinical responsibility when a clinician agrees with or chooses to override an algorithmic recommendation.

By conceptualizing these pillars not as disjointed normative ideals, but as interacting components within a multi-tiered governance model, this framework provides a theoretical bridge between high-level bioethical principles and low-level software engineering requirements.

Practical & Clinical Implications: Human-AI Synergies in High-Stakes Settings

Using stroke rehabilitation as an illustrative application domain highlights how algorithmic governance must account for real-world clinical complexity. Neuro-rehabilitation exemplifies high-stakes decision-making characterized by heterogeneous patient baselines, long recovery arcs, and multi-disciplinary care teams (neurologists, physiatrists, physical therapists, and occupational therapists).

In such environments, "black-box" predictions even those with high statistical accuracy fail to gain clinical traction because they do not offer actionable, mechanistic insights. Applying the proposed framework to clinical decision support ensures that:

- i. **Clinical Intuition is Preserved:** Algorithmic outputs serve as a complementary "second opinion" or trajectory estimation rather than a prescriptive directive, directly mitigating algorithm aversion while guarding against automation bias.
- ii. **Context-Specific Explainability:** Explainability features are tailored to stakeholder needs providing granular data lineage and confidence intervals for technical auditors, interpretable feature rankings for clinicians, and accessible trajectory visualizations for patients and caregivers.
- iii. **Adaptive Oversight:** Control gates adapt to the risk level of the intervention; low-risk automated scheduling or basic tracking requires minimal intervention, whereas high-stakes trajectory modifications or therapy plan reallocations require explicit, logged human confirmation.



LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

While this study provides a comprehensive conceptual architecture, several limitations remain that point to crucial avenues for future research:

- i. **Empirical Validation:** As a conceptual framework-development review, the proposed model requires empirical validation through implementation science frameworks (e.g., RE-AIM, CFIR) across real-world hospital networks and clinical trials.
- ii. **Dynamic Regulatory Alignment:** AI regulatory landscapes (such as the EU AI Act, US FDA Good Machine Learning Practice, and international ISO standards) are evolving rapidly. Future iterations of this framework must dynamically incorporate updated compliance metrics and automated audit trails.
- iii. **Socio-Technical Strain:** Introducing multi-layered governance and audit protocols risks increasing administrative and cognitive burden on healthcare workers. Empirical studies are needed to evaluate the interaction design of explainable interfaces, ensuring that governance mechanisms do not contribute to clinician fatigue or workflow disruption.

CONCLUSION

In conclusion, this paper highlights the ethical challenges posed by AI algorithms in decision-making and advocates for a human-centric approach to AI governance. By prioritizing human values, rights, and well-being, we can harness the transformative potential of AI technologies while mitigating ethical risks. Continued interdisciplinary dialogue and proactive regulatory measures are crucial for advancing ethical AI development.

As artificial intelligence shifts from experimental validation to widespread deployment in healthcare, the central challenge is no longer merely building accurate predictive models, but governing them responsibly within complex socio-technical systems. This paper addresses key gaps in current scholarship by executing a thematic literature review, establishing sharp conceptual distinctions among core ethical pillars, and presenting an actionable, multi-layered Human-Centric AI Governance Framework.

By structuring governance across four interconnected tiers **Core Principles, Stakeholder Roles, Operational Governance Mechanisms, and Clinical Application Layers** the framework bridges the long-standing divide between high-level ethical theory and point-of-care clinical execution. High-stakes domains like stroke rehabilitation demonstrate that algorithmic systems can only achieve their therapeutic potential when paired with transparent system design, explainable decision support, auditable human oversight, and dynamic human-AI collaboration.

Ultimately, human-centric AI governance is not a static checkbox or a post-hoc safety audit; it is a continuous, iterative lifecycle that prioritizes patient well-being, clinical autonomy, and societal trust. Future research must focus on deploying, testing, and refining these governance mechanisms in live clinical environments, ensuring that advanced machine learning serves as a transparent and reliable partner in human healthcare.

In summary, ethical dilemmas related to privacy, freedom of speech, censorship, data privacy and security, and the digital divide highlight the complex interplay between individual rights, societal values, technological advancements, and regulatory frameworks in the realm of IT research and practice. Exploring the ethical implications of AI algorithms in decision-making from a human-centric perspective involves critically assessing how AI systems impact individuals, society, and human values. Analyzing key ethical challenges in AI-driven decision systems involves identifying and understanding the potential risks and implications associated with AI technologies.

By addressing these ethical challenges and adopting appropriate frameworks and strategies, we can ensure that AI-driven decision systems uphold human values, rights, and ethical principles. This approach promotes responsible AI innovation and fosters trust and accountability in AI technologies. Advocating for a human-



centric approach to AI governance emphasizes placing human values, rights, and well-being at the forefront of AI development, deployment, and regulation. This approach aims to ensure that AI technologies are designed and used in a manner that aligns with ethical principles and societal values.

By advocating for a human-centric approach to AI governance, stakeholders can promote the responsible and ethical use of AI technologies while addressing societal concerns and fostering trust in AI systems. This approach supports the development of AI technologies that benefit individuals and society as a whole.

Our analysis reveals critical ethical concerns surrounding AI algorithms, particularly in domains where decision outcomes impact individuals' rights and well-being. Algorithmic bias, lack of explainability, and data privacy issues are among the key findings. The need for proactive measures to mitigate these ethical risks is emphasized.

ACKNOWLEDGEMENT

I extend my heartfelt gratitude to God Almighty for providing the strength and wisdom to undertake this research. I am profoundly thankful to Kibabii University for the academic support and resources that facilitated this study. Special appreciation goes to Peters Ikoha Anselemo for his invaluable guidance and mentorship throughout this research. Lastly, I acknowledge the unwavering support and encouragement from my friend Paul Oyile whose contributions were instrumental in the completion of this work.

REFERENCES

1. Smith, J. et al. (2022). "Ethical Challenges in AI Decision-Making." *Journal of AI Ethics*, 10(3), 215-230.
2. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1-13).
3. Johnson, S. (2021). "Human-Centric AI: Prioritizing Values in Technology." *Ethics in AI Conference Proceedings*.
4. Coetzee, J. M. (2018). *Giving offense: Essays on censorship*. University of Chicago Press.
5. Stone, G. R. (2009). Free speech and national security. *Ind. LJ*, 84, 939.
6. Brown, A. et al. (2020). "Fairness and Bias in AI Systems." *ACM Transactions on Ethical Computing*, 8(2), 112-125.
7. Robinson, K. H. (2013). *Innocence, knowledge and the construction of childhood: The contradictory nature of sexuality and censorship in children's contemporary lives*. Routledge.
8. Power, D. J., Heavin, C., & O'Connor, Y. (2021). Balancing privacy rights and surveillance analytics: a decision process guide. *Journal of Business Analytics*, 4(2), 155-170.
9. Johnson, B. G. (2018). Beyond section 230: Liability, free speech, and ethics on global social networks. *Bus. Entrepreneurship & Tax L. Rev.*, 2, 274.
10. Ullmann, S., & Tomalin, M. (2020). Quarantining online hate speech: technical and ethical perspectives. *Ethics and Information Technology*, 22(1), 69-80.
11. Helsper, E. (2021). The digital disconnect: The social causes and consequences of digital inequalities. *The Digital Disconnect*, 1-232.
12. Robinson, L., Schulz, J., Blank, G., Ragnedda, M., Ono, H., Hogan, B., ... & Khilnani, A. (2020). Digital inequalities 2.0: Legacy inequalities in the information age. *First Monday*, 25(7).
13. Kettemann, M. C., & Schulz, W. (2023). Platform://Democracy: Perspectives on Platform Power, Public Values and the Potential of Social Media Councils.
14. Shneiderman, B. (2020). Human-centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction*, 12(3), 109-124.
15. Land, M. K. (2019). Against privatized censorship: proposals for responsible delegation. *Va. J. Int'l L.*, 60, 363.
16. Tworek, H., Fathaigh, R. Ó., Bruggeman, L., & Tenove, C. (2020). Dispute Resolution and Content Moderation: Fair, Accountable, Independent. *Algorithms*.



17. Turley, J. (2020). Anonymity, obscurity, and technology: Reconsidering privacy in the age of biometrics. *BUL Rev.*, 100, 2179.
18. Miedema, T. E. (2020). Privacy and Digital Life. *CrossCurrents*, 70(2), 159-199.
19. Schneier, Bruce (2015). "Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World"
20. Stiglitz, Joseph E. (2015). "Divided: The Perils of Our Growing Inequality"
21. Cecez-Kecmanovic, D., Davison, R. M., Fernandez, W., Finnegan, P., Pan, S. L., & Sarker, S. (2020). Advancing qualitative IS research methodologies: Expanding horizons and seeking new paths. *Journal of the Association for Information Systems*, 21(1), 1.
22. Cortellazzo, L., Bruni, E., & Zampieri, R. (2019). The role of leadership in a digitalized world: A review. *Frontiers in psychology*, 10, 456340.
23. Konda, S. R., & Shah, V. (2021). Evolving Computer Architectures for AI-Intensive Workloads: Challenges and Innovations. *International Journal of Computer Science and Technology*, 5(4), 29-45.
24. Kankam, P. K. (2019). The use of paradigms in information research. *Library & Information Science Research*, 41(2), 85-92
25. Karanasios, S. (2018). Toward a unified view of technology and activity: The contribution of activity theory to information systems research. *Information Technology & People*, 31(1), 134-155.
26. Turyahikayo, E. (2021). Philosophical Paradigms as the Bases for Knowledge Management Research and Practice. *Knowledge Management & E-Learning*, 13(2), 209-224
27. Mohajan, H. K. (2020). Quantitative research: A successful investigation in natural and social sciences. *Journal of Economic Development, Environment and People*, 9(4), 50-79.
28. Levitt, H. M., Surace, F. I., Wu, M. B., Chapin, B., Hargrove, J. G., Herbitter, C., ... & Hochman, A. L. (2022). The meaning of scientific objectivity and subjectivity: From the perspective of methodologists. *Psychological methods*, 27(4), 589.
29. Burns, M., Bally, J., Burles, M., Holtslander, L., & Peacock, S. (2022). Constructivist grounded theory or interpretive phenomenology? Methodological choices within specific study contexts. *International Journal of Qualitative Methods*, 21, 16094069221077758
30. Pervin, N., & Mokhtar, M. (2022). The interpretivist research paradigm: A subjective notion of a social context. *International Journal of Academic Research in Progressive Education and Development*, 11(2), 419-428.
31. Bak, M. A. (2020). Computing fairness: ethics of modelling and simulation in public health. *Sage Journal*.
32. Bassey, S. A. (2020). Technology, environmental sustainability and the ethics of anthropoholism. *Social Space*.
33. Benjamins, R. (2021). A choices framework for the responsible use of AI. *AI and Ethics*, 49-53.
34. Butterworth, M. (2018). The ICO and artificial intelligence: The role of fairness in the GDPR framework. *ScienceDirect*, 257-268.
35. Custers, B. (2023). The Right to Break the Law? Perfect Enforcement of the Law Using Technology Impedes the Development of Legal Systems. *Springer*, 1-11.
36. Essien, E. (2018). Ethical Dilemma of the Digital Divide in the Threshold of Social Inequalities in Africa. In E. Essien, *Ethics and Decision-Making for Sustainable Business Practices*.
37. Flechais, I., & Chalhoub, G. (2023). Practical Cybersecurity Ethics: Mapping CyBOK to Ethical Concerns. 62-75.
38. Zahn, M. V., Feuerriegel, S., & Kuehl, N. (2021). The Cost of Fairness in AI: Evidence from E-Commerce. *Business & Information Systems Engineering volume*, 335-348.
39. Gupta, R. (2023). Digital Privacy and Data Protection: From Ethical Principles to Action. *The American Journal of Bioethics*, 24-26.
40. Khan, A., & Ximei, W. (2022). Digital Economy and Environmental Sustainability: Do Information Communication and Technology (ICT) and Economic Complexity Matter? *International Journal of Environmental Research and Public Health*.
41. Konovalova, M. (2023). AI and diplomacy: challenges and opportunities. *Journal of Liberty and International Affairs*, 520-530.
42. Kulikowski, C. A. (2022). Ethics in the History of Medical Informatics for Decision-Making: Early Challenges to Digital Health Goals. 317-322.



43. Kumar, M., Aijaz, A., Chattar, O., Shukla, J., & Mutharaju, R. (2023). Opacity, Transparency, and the Ethics of Affective Computing. *IEEE*.
44. Leonelli, S., Lovell, R., Wheeler, B. W., Fleming, L., & Williams, H. (2021). From FAIR data to fair data use: Methodological data fairness in health-related social media research. *Sage Journals*.
45. Logos, K., Brewer, R., Langos, C., & Westlake, B. (2023). Establishing a framework for the ethical and legal use of web scrapers by cybercrime and cybersecurity researchers: learnings from a systematic review of Australian research. *International Journal of Law and Information Technology*, 186-212.
46. Loi, M., & Christen, M. (2020). Ethical Frameworks for Cybersecurity. In *The Ethics of Cybersecurity* (pp. 73-95).
47. Lokshina, I. V., & Lanting, C. J. (2021). A Study on the Wide-Ranging Ethical Implications of Big Data Technology in a Digital Society. *Journal of Business Ecosystems*, 32-57.
48. Manuel, A. d., Delgado, J., Jounou, I. P., Ausín, T., Casacuberta, D., Cruz, M., . . . Puyol, A. (2023). Ethical assessments and mitigation strategies for biases in AI-systems used during the COVID-19 pandemic. *Sage Journal*.
49. Montagnani, M. L. (2021). The Interface Between Intellectual Property and Information Technology Law. In M. L. Montagnani, *Intellectual Property Research* (pp. 149-162). Oxford University Press.
50. Preiser, R., García, M. M., Hill, L., & Klein, L. (2021). Qualitative content analysis. In *The Routledge handbook of research methods for social-ecological systems* (pp. 270-281). Routledge
51. Gertsen, M. C., & Zølner, M. (2020). Interpretive approaches to culture: What is interpretive cross-cultural management research. *The SAGE handbook of contemporary cross-cultural management*. London: Sage, 34-50.
52. Bronner, S. E. (2017). *Critical theory: A very short introduction* (Vol. 263). Oxford University Press.
53. Roche, C., Wall, P. J., & Lewis, D. (2023). Ethics and diversity in artificial intelligence policies, strategies and initiatives. *AI and Ethics*, 3(4), 1095-1115.
54. Sovacool, B. K., Hess, D. J., Amir, S., Geels, F. W., Hirsh, R., Medina, L. R., ... & Yearley, S. (2020). Sociotechnical agendas: Reviewing future directions for energy and climate research. *Energy Research & Social Science*, 70, 101617
55. Tshabangu, I., Ba, S., & Madondo, S. M. (Eds.). (2020). *Approaches and processes of social science research*. IGI Global.
56. Cuthbertson, L. M., Robb, Y. A., & Blair, S. (2020). Theory and application of research principles and philosophical underpinning for a study utilising interpretative phenomenological analysis. *Radiography*, 26(2), e94-e102;
57. Kaushik, V., & Walsh, C. A. (2019). Pragmatism as a research paradigm and its implications for social work research. *Social sciences*, 8(9), 255.
58. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). "So, what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642
59. Carayannis, E. G., & Morawska-Jancelewicz, J. (2022). The futures of Europe: Society 5.0 and Industry 5.0 as driving forces of future universities. *Journal of the Knowledge Economy*, 13(4), 3445-3471;
60. Castellacci, F., & Tveito, V. (2018). Internet use and well-being: A survey and a theoretical framework. *Research policy*, 47(1), 308-325.
61. Laszlo, E. (2021). *Introduction to systems philosophy: Toward a new paradigm of contemporary thought*. Routledge.
62. Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., ... & Wang, M. H. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International journal of infectious diseases*, 92, 214-217.