Ethics and Accountability Frameworks for Generative AI Systems

[1]Mr.Sidharth Sharma

[1]*Vice President – IT Projects/Audits, JP Morgan Chase. Inc, 545 Washington Blvd Jersey City, NJ 07310 – US.*

[1]Corresponding Author's email: *infosidharthsharma@gmail.com*

Abstract. Intense discussions concerning the hazards and ethical ramifications of generative artificial intelligence were sparked by its introduction and broad societal adoption. Traditional discriminative machine learning carries hazards that are frequently different from these risks. A scoping review on the ethics of generative artificial intelligence, with a focus on big language models and text-to-image models, was carried out in order to compile the recent discourse and map its normative notions. Enforcing accountability, responsibility, and adherence to moral and legal standards will become more challenging as artificial intelligence systems get more adept at making decisions on their own. Here, a user-centered, realism-inspired method is suggested to close the gap between abstract concepts and routine research procedures. It lists five particular objectives for the moral application of AI: 1) comprehending model training and output, including bias mitigation techniques; 2) protecting copyright, privacy, and secrecy; 3) avoiding plagiarism and policy infractions; 4) applying AI in a way that is advantageous over alternatives; and 5) employing AI in a transparent and repeatable manner. Every objective is supported by workable plans, real-world examples of abuse, and remedial actions. This paper will discuss the nature of an accountability framework and related concerns in order to enable the organized responsibility for assignment and proof of AI systems. The suggested architecture for regulating AI incorporates crucial components like transparency, human oversight, and adaptability to address the issues with accountability that have been brought to light. Some crucial suggestions for putting the framework into practice and growing it were also provided through industrial case studies, guaranteeing that companies increase compliance, trust, and responsible AI technology adoption.

Keywords. Artificial intelligence, Accountability, Generative AI, Ethics

1. INTRODUCTION

Intelligent self-governing decision-making systems are relatively new, yet given their many benefits and wider ramifications, they may be considered the greatest achievement humanity has ever made. The issue of decision-making without human interaction is extremely concerning as these systems permeate several industries, such as insurance and healthcare. The intricacy of AI's operations makes it difficult to explain the decisions made by the algorithms, which leads to the problem of assigning blame when an error or bias occurs. AI may encounter moral dilemmas, legal challenges, and a decline in public trust as a result of this lack of transparency. This is a result of our society's growing reliance on AI to make crucial judgments. Therefore, it is necessary to provide an appropriate framework to direct AI's behavior so that it complies with legal requirements and societal norms. Accountability frameworks provide a framework for resolving AI decision-making risks in a fair, open, and rational way. Although the academic community benefits greatly from these meta-studies, there isn't one that specifically addresses the collection of moral dilemmas related to the most recent explosion in generative AI applications.

This study, a scoping review, aims to bridge this knowledge gap and offer a useful summary for academics, AI professionals, decision-makers, journalists, and other pertinent parties. We synthesize the discourse's specifics, map normative notions, address imbalances, extract the body of knowledge on the ethics of generative AI, and supply a foundation for future research and technological governance using a methodical literature search and coding technique. A diverse dataset comprising both human and AI-generated text samples, we demonstrate the superiority of our method in accurately discerning between the two. By advancing AI-generated text detection techniques, our research seeks to mitigate the risks associated with the proliferation of AI generated content and foster trust in digital communication platforms.

2. LITERATURE SURVEY

According to Novelli et al. (2023), the issue of accountability in artificial intelligence (AI) is both contentious and ambiguous, making it challenging to define. This concept can be understood through the lens of answerability, which encompasses the recognition of authority, the capacity for interrogation, and the limitation of power. Consequently, accountability is crucial in the governance of AI, particularly as these systems are expected to undertake decision-making roles. It is a fundamental principle enshrined in both the AI Act and the General Data Protection Regulation (GDPR) in Europe. However, despite the existence of these principles, there remains a lack of clarity regarding their implementation across various systems. The inquiry into accountability within the sociotechnical framework of AI systems raises numerous complex issues. These systems comprise both human and technological elements, blurring the lines of responsibility for the outcomes of AI-driven decisions. The inherent characteristics of artificial intelligence algorithms—specifically their informal structure and non-deterministic nature—complicate the assignment of clear responsibility and the transparency of decision-making processes. In relation to the current discourse on accountability in AI as outlined in key European regulations, it is noteworthy that the predominant emphasis tends to be on compliance and oversight. This focus often overshadows other critical aspects, such as transparency and ethical considerations. There is a notable absence of a well-defined and systematically developed theory of accountability that explicitly delineates the various types and levels of enforcement, as well as the sociotechnical dimensions of AI, a gap that has yet to be addressed by scholars and policymakers.

As per what Verdiesen et al 2020 has to say, one of the prominent issues dealing with ethical and legal aspects of autonomous systems and governance is accountability. accountability in the AWS context is blurred. In most cases, responsibility is perceived retrospectively, where the relevant actors make justifications of their actions and after the event takes place. With reference to AWS, this raises crucial issues surrounding the levels of so-called ‘accountability gaps,’ under which nobody is deemed liable. This is made more complicated by the fact that such systems are often complex and fully autonomous, which at times makes linking the decision- making process back to the operators difficult, if not impossible. Literature confirms how important accountability is in autonomous systems, and control mechanisms that are said to assist in enabling it. Some characteristics of contemporary tactical missions management that are, inter alia, mission mandate and rules of engagement may be irrelevant or insufficient for fully autonomous systems. They should therefore propose a Comprehensive Human Oversight Framework which covers technical and socio-technical systems alongside governance. Although these shortcomings have not entirely destroyed earlier attempts to construct responsibilities and accountability in autonomous systems, there are still significant issues with the applicable governance and oversight mechanisms in the deployment phase of autonomous systems. The paper argues for a more comprehensive strategy that takes into account the challenges brought about by AWS's independence. Without these steps, there's a danger that accountability won't be sufficient, which could lead to ethical and legal issues.

Yazdanpanah et al. 2023 explains that the incorporation of autonomous systems into society raises new challenges surrounding accountability, such as determining how these systems should operate within legal and ethical boundaries. Responsibility regarding autonomous systems technology IS considers the ability to allocate blame in two contexts: anticipatory where a person can be held to account before an actual event occurs, and retrospective where the true event can be assessed. Do note that the ability to numerically measure and allocate blame shows the considerable gap in accountability, especially when using automated systems that self-manage a large volume of operations. The literature points out that there is the crucial task of creating a proper framework that could enable autonomous systems to make ad satisfying decisions regarding responsibility to guarantee the reliability of these systems and their compliance with legal and ethic norms. Implementing systems intended to foster cooperation is a requirement that governs self-sufficient systems and their implications on society, and the more dire the consequences of failure, the deeper this logic is rooted.

3. PROPOSED SYSTEM

Accountability Frameworks for AI Decision-Making Systems

To ensure accountability for AI decision-making systems, it is crucial to have access to high-quality and comprehensive data. This involves collecting a wide range of first-hand information about the functioning of existing AI systems, their decision-making patterns, the outcomes of their actions, and the interactions between users and the AI systems. This data collection should span various industries where AI is applied, such as healthcare, finance, and automotive.

Additionally, information should be gathered from relevant authorities and organizations that regulate the use of AI technologies. The compiled data should cover a spectrum of cases, from instances where the AI systems performed well to those with negative consequences for different segments of society. Particular emphasis should be placed on understanding how accountability was addressed or lacking in certain incidents.

By gathering this diverse set of data, accountability frameworks can be developed to ensure that AI decision-making systems are transparent, responsible, and aligned with societal well-being.

The Designing of the Machine Learning Models Outline Forensic reoccurrence examines that ML models are vital in ascertaining the functions of AI systems, particularly when self-governance and accountability matters. These models are structured by the fact that these are decision-making models intending to replicate real life challenges in AI application (Falco et al, 2021). Then, Strong accountability models are further analyzed to determine where there are gaps and where positive patterns exist within are supported by ML models.

The interpretability feature of the models ensures that the decisions produced by the AI models can be comprehended by the human mind (Percy et al., 2021). Because it enables the analysis of the decision-making process of the developed AI systems, this is especially significant in terms of compliance with the accountability frameworks. In addition to the other model explainability methods, Mlda incorporates feature importance analysis to give this transparency. Additionally, given the speed at which technology is developing, the models are designed to be trainable in a way that permits continuous training with fresh data.

Implementation and Deployment

The delivery of accountability frameworks and associated ML models is done in phases. Their frameworks and models are initially tested under controlled settings to demonstrate their validity. This final phase encompasses practicing in different types of industries to verify the capability of the models to evaluate and optimize accountability in the respective fields (Laitinen and Sahlgren 2021). Any deficiencies that might be observed are fed back into the models and frameworks to fine tune it.

Pilot schemes in selected industries are initiated first after successful test. It, however, is closely monitored with regular collection and analysis of data to evaluate the effectiveness of these frameworks in promoting accountability (Bjørlo et al. 2021). Stakeholders' training, which include the developers of AI and the regulators, is also conducted at this phase so that they know how to utilize the laid down frameworks and interpret the outputs of the ML models.

4. RESULTS AND DISCUSSION

Evaluation of the Framework for the Increase in the Responsibility

Consequently, when applied to autonomous AI decision-making systems, accountability with the proposed framework has the potential to improve the overall transparent functioning of AI. According to Loi and Spielkamp (2021), the framework's explanation revealed that by integrating the aforementioned measures, AI systems could offer more transparent decision trails that make it simple to identify the appropriate parties when decisions had unfavorable effects. Transparency and human supervision, two fundamental tenets of the architecture that was provided, were also essential in the case of accountability. These two principles greatly reduced the main disadvantage of autonomous AI, which was the absence of operational transparency.

Examples of cases and results of applying the criteria

In the case studies carried out across various economic sectors, the framework demonstrated both flexibility and stability. For example, within the healthcare sector, it facilitated clearer delineation of responsibilities as AI systems became involved in diagnosis and treatment planning. In the finance sector, it played a crucial role in identifying decision-making processes within automated trading systems, thereby aiding in regulatory compliance (de Almeida et al. 2021). The framework's alignment with existing regulations governing AI usage further ensured that AI systems adhered to legal standards, preventing violations and raising awareness among all relevant stakeholders regarding their obligations. The findings suggest that the proposed framework not only enhances accountability but also fosters greater confidence in the deployment of autonomous AI systems across various industries.

The implementation of the proposed accountability framework and its auditing processes has demonstrated significant advancements in enhancing the autonomy of AI decision-making systems by improving their accuracy. By addressing national responsibilities and incorporating human moderators, this framework facilitates the ethical integration of AI across various sectors. Although the framework encounters new challenges due to technological progress, it remains committed to upholding core principles such as transparency in decision-making and comprehensive decision management (Taeihagh, 2021). The addition of real-time monitoring capabilities and adherence to a continuous learning approach could further enhance the framework. Ultimately, the current research underscores the importance of accountability in fostering trust and ensuring the ethical application of AI within society.

It is recommended that future research and development efforts focus on enhancing the proposed accountability framework, as the challenges related to the implementation of automated decision-making systems in artificial intelligence are expected to evolve over time. Consequently, the decision-making bot may need to adapt in accordance with advancements in various AI technologies to fulfill the framework's requirements (Malgieri and Pasquale 2022). A notable avenue for future research involves the incorporation of real-time monitoring technologies and adaptive learning, which could elevate the practical application of the proposed framework, thereby improving accountability management in complex and dynamic system environments. Additionally, a comprehensive examination of the interplay between AI ethics and legal regulations on an international scale may provide valuable insights for establishing widely accepted accountability standards. Collaboration among the developers of AI systems, ethicists, and policymakers will be crucial in facilitating these necessary changes.

5. CONCLUSION

The implementation and maintenance of has-independent-accountability mechanisms are necessary to ensure that autonomous AI decision-making systems continue to function as planned. The proposed framework has demonstrated potential in increasing transparency, bridging the accountability gap, and guaranteeing that AI systems adhere to legal and ethical standards, all of which are in line with the objectives of the study. Because it outlines the roles and accountability, the framework's adoption gives many industries peace of mind that AI technology will be closely watched. However, these frameworks will need to be further improved and refined in the future as AI and its applications become more sophisticated.

REFERENCES

1. Jasper Gnana Chandran, J., Karthick, R., Rajagopal, R., & Meenalochini, P. (2023). Dual-channel capsule generative adversarial network optimized with golden eagle optimization for pediatric bone age assessment from hand X-ray image. *International Journal of Pattern Recognition and Artificial Intelligence*, *37*(02), 2354001.
2. Karthick, R., Prabha, M., Sabapathy, S. R., Jiju, D., & Selvan, R. S. (2023, October). Inspired by social-spider behavior for microwave filter optimization, swarm optimization algorithm. In *2023 International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS)* (Vol. 1, pp. 1-4). IEEE.
3. Vijayalakshmi, S., Sivaraman, P. R., Karthick, R., & Ali, A. N. (2020, September). Implementation of a new Bi-Directional Switch multilevel Inverter for the reduction of harmonics. In *IOP Conference Series: Materials Science and Engineering* (Vol. 937, No. 1, p. 012026). IOP Publishing.
4. Kiruthiga, B., Karthick, R., Manju, I., & Kondreddi, K. (2024). Optimizing harmonic mitigation for smooth integration of renewable energy: A novel approach using atomic orbital search and feedback artificial tree control. *Protection and Control of Modern Power Systems*, *9*(4), 160-176.
5. Sulthan Alikhan, J., Miruna Joe Amali, S., & Karthick, R. (2024). Deep Siamese domain adaptation convolutional neural network-based quaternion fractional order Meixner moments fostered big data analytical method for enhancing cloud data security. *Network: Computation in Neural Systems*, 1-28.
6. Sakthi, P., Bhavani, R., Arulselvam, D., Karthick, R., Selvakumar, S., & Sudhakar, M. (2022, September). Energy efficient cluster head selection and routing protocol for WSN. In *AIP Conference Proceedings* (Vol. 2518, No. 1). AIP Publishing.
7. Aravindaguru, I., Arulselvam, D., Kanagavalli, N., Ramkumar, V., & Karthick, R. (2022, September). Space cloud in cubesat-Consigning expert system to space. In *AIP Conference Proceedings* (Vol. 2518, No. 1). AIP Publishing.
8. Karthick, R., Prabaharan, A. M., & Selvaprasanth, P. (2019). A Dumb-Bell shaped damper with magnetic absorber using ferrofluids. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*.
9. Selvan, R. S., Wahidabanu, R. S. D., Karthick, B., Sriram, M., & Karthick, R. (2020). Development of Secure Transport System Using VANET. *TEM (H-Index)*, *82*.
10. Karthick, R., & Sundararajan, M. (2018). Optimization of MIMO Channels Using an Adaptive LPC Method. *International Journal of Pure and Applied Mathematics*, *118*(10), 131-135.
11. Lopez, S., Sarada, V., Praveen, R. V. S., Pandey, A., Khuntia, M., & Haralayya, D. B. (2024). Artificial intelligence challenges and role for sustainable education in india: Problems and prospects. *Sandeep Lopez, Vani Sarada, RVS Praveen, Anita Pandey, Monalisa Khuntia, Bhadrappa Haralayya (2024) Artificial Intelligence Challenges and Role for Sustainable Education in India: Problems and Prospects. Library Progress International*, *44*(3), 18261-18271.
12. Kumar, N., Kurkute, S. L., Kalpana, V., Karuppannan, A., Praveen, R. V. S., & Mishra, S. (2024, August). Modelling and Evaluation of Li-ion Battery Performance Based on the Electric Vehicle Tiled Tests using Kalman Filter-GBDT Approach. In *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)* (pp. 1-6). IEEE.
13. Sharma, S., Vij, S., Praveen, R. V. S., Srinivasan, S., Yadav, D. K., & VS, R. K. (2024, October). Stress Prediction in Higher Education Students Using Psychometric Assessments and AOA-CNN-XGBoost Models. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1631-1636). IEEE.
14. Yamuna, V., Praveen, R. V. S., Sathya, R., Dhivva, M., Lidiya, R., & Sowmiya, P. (2024, October). Integrating AI for Improved Brain Tumor Detection and Classification. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1603-1609). IEEE.
15. Anuprathibha, T., Praveen, R. V. S., Jayanth, H., Sukumar, P., Suganthi, G., & Ravichandran, T. (2024, October). Enhancing Fake Review Detection: A Hierarchical Graph Attention Network Approach Using

Text and Ratings. In *2024 Global Conference on Communications and Information Technologies (GCCIT)* (pp. 1-5). IEEE.