Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 3
March 2024

A Comprehensive Study on Generative AI Risks in the Financial Sector: Challenges and Solutions

Danish Kapoor[1*] , Anmol Bhatia[1]

[1]Department of Multidisciplinary Engineering, The NorthCap University, Gurugram, Haryana, India.

Abstract

Generative AI systems have revolutionized content creation but also pose ethical, societal, and security risks. This paper examines these risks and proposes mitigation strategies. A comprehensive AI risk assessment study was conducted in the finance sector in which questionnaires were administered to 50 key stakeholders, including executives, risk managers, data scientists, and compliance officers. The analysis revealed significant concerns, including biased decision-making, security vulnerabilities, and challenges in regulatory compliance. Emphasis was placed on the need for rigorous testing and monitoring to mitigate bias, robust security measures against cyber threats, and proactive engagement with regulators. The study underscores the multifaceted nature of AI risks in finance and highlights the importance of a holistic approach to risk management. The findings suggest prioritizing fairness, transparency, and accountability in AI deployment, implementing robust security measures, staying abreast of evolving regulations, and enhancing risk governance structures. This study provides valuable insights for financial institutions to navigate the complexities of AI risks and ensure responsible and ethical AI deployment.

Keywords: *Generative AI, Risk Assessment, Finance Sector, Mitigation Strategies, and Ethical Deployment.*

Introduction

Generative Artificial Intelligence (AI) systems have emerged as transformative tools across various sectors, facilitating the creation of novel content with unprecedented efficiency and scale. However, alongside their potential benefits, these systems also introduce complex risks, ranging from the propagation of biased or malicious outputs to the proliferation of deepfakes and misinformation [1]. Generative AI systems, underpinned by sophisticated deep learning algorithms, have revolutionized content creation by autonomously generating diverse outputs encompassing images, text, and music [2]. While these systems hold tremendous promise for innovation and creativity, they also pose significant risks that necessitate careful consideration and management.

Numerous research papers have delved into various aspects of generative AI and its associated risks. One of the pioneering works in generative AI is the introduction of Generative Adversarial Networks (GANs) by Goodfellow et al. [3]. GANs consist of two neural networks, a generator, and a discriminator, engaged in a competitive game where the generator aims to produce realistic outputs while the discriminator seeks to distinguish between real and generated samples. This seminal work laid the foundation for subsequent research in generative modeling and adversarial training. Furthermore, Radford et al. [4] presented a significant advancement with the introduction of GPT (Generative Pre-trained Transformer), a state-of-the-art language model capable of generating coherent and contextually relevant text. This breakthrough in natural language generation has far-reaching implications, including applications in automated writing, conversational agents, and content generation platforms.

The proliferation of generative AI raises profound ethical concerns, particularly regarding the creation and dissemination of synthetic content. Research by Schwartz et al. [5] emphasized the importance of ethical considerations in the development and deployment of generative AI systems. They advocated for transparency, accountability, and user consent to mitigate potential harms associated with AI-generated content, such as misinformation, manipulation, and infringement of privacy rights.

Additionally, Liang et al. [6] investigated the legal implications of generative AI-generated content, highlighting challenges related to intellectual

property rights, copyright infringement, and liability. The study underscored the need for updated legal frameworks to address the unique challenges posed by AI-generated content and ensure fair compensation and protection for content creators and rights holders.

Generative AI systems also pose significant security and privacy risks, particularly in sensitive domains such as healthcare and finance. Wang & Zhang [7] explored the privacy implications of generative AI in healthcare, emphasizing the need for robust security measures to safeguard sensitive patient data. As AI technologies are increasingly integrated into healthcare systems for tasks like medical imaging analysis and patient diagnostics, concerns regarding data privacy and security become paramount.

Moreover, recent research by khazraji et al. [8] investigated the potential misuse of generative AI for creating deepfake videos, highlighting the implications for privacy, security, and disinformation campaigns. The study called for robust detection and mitigation strategies to combat the proliferation of deepfake content in online platforms and social media.

The societal implications of generative AI extend beyond privacy and security concerns to encompass broader cultural, artistic, and identity-related considerations. Oksanen et al. [9] examined the societal implications of AI-generated content, including its impact on cultural heritage, artistic expression, and identity. The study emphasized the need for interdisciplinary collaboration and stakeholder engagement to address the ethical, legal, and societal challenges posed by generative AI technologies.

Despite the significant advancements in generative AI, several challenges remain unresolved. Future research efforts should focus on developing robust risk assessment frameworks tailored to the unique characteristics of generative AI systems. Additionally, interdisciplinary collaboration between AI researchers, ethicists, policymakers, and stakeholders is essential to ensure responsible development and deployment of generative AI technologies.

Recent literature highlights the growing concern regarding the risks posed by Generative AI within the finance sector. For instance, Smith et al. [10] examined the implications of Generative AI on financial fraud detection, emphasizing the challenges associated with distinguishing between genuine and fraudulent activities. Similarly, Jones and Lee [11] conducted a study on the vulnerability of AI-driven trading algorithms to manipulation and exploitation, underscoring the need for robust regulatory frameworks to mitigate market risks. Furthermore, Chen and Wang [12] investigated the ethical considerations surrounding the use of Generative AI in credit scoring, revealing concerns related to bias and fairness in decision-making processes. In addition, Gupta et al. [13] explored the security vulnerabilities of AI-powered chatbots in customer service, highlighting the risks of data breaches and privacy infringements. Conversely, Kim et al. [14] focused on the potential societal implications of Generative AI in personal finance management, indicating the need for transparent and accountable AI systems to maintain consumer trust. Moreover, Li and Zhang [15] examined the regulatory challenges associated with the adoption of Generative AI in compliance monitoring, suggesting the necessity for harmonized regulatory standards to address cross-border complexities. Furthermore, Wu and Liu [16] conducted a study on the legal liabilities arising from AI-driven investment recommendations, emphasizing the importance of clear accountability frameworks to mitigate legal risks. Additionally, Park and Choi [17] analyzed the impact of Generative AI on financial stability, highlighting the need for systemic risk assessment mechanisms to safeguard against market disruptions. Wang and Huang [18] investigated the role of Generative AI in asset pricing models, revealing challenges related to model interpretability and reliability. Despite these endeavors, substantial gaps persist in comprehending the enduring consequences of Generative AI on financial markets and the effectiveness of risk reduction approaches. Additional research is required to investigate the interaction among technological progressions, regulatory structures, and ethical deliberations in molding the future landscape of AI-driven finance.

The main goal of this research is to explore the diverse risks linked with incorporating advanced technological instruments in financial activities. Through an extensive examination, the study

intends to illuminate potential challenges like partial decision-making, breaches of data, and failure to adhere to regulations. The research aims to fill gaps in research by providing understanding into the complex interaction among technology, ethics, and societal influence within the domain of generative AI in the financial sector.

2. Understanding Generative AI Risks

Generative Artificial Intelligence (AI) systems mark a revolutionary progress in technology, allowing for the independent production of varied content across multiple fields. Nonetheless, this advancement also presents a plethora of dangers that require careful examination and control. Grasping the hazards linked with generative AI is crucial for effectively navigating its intricacies. This segment explores the ethical, societal, and security issues inherent in generative AI systems, clarifying their consequences and hurdles.

2.1 Ethical Concerns

Generative AI systems confront ethical dilemmas concerning the integrity and fairness of the content they produce. The reliance on training data imbued with biases or inaccuracies can lead to the perpetuation of stereotypes or dissemination of misinformation. For instance, language models trained on internet text may inadvertently generate discriminatory or harmful content due to underlying biases [19]. The ethical implications of such outputs extend to their impact on individuals and society, necessitating ethical frameworks and guidelines to mitigate potential harms.

Recent studies highlight the significance of addressing ethical concerns in generative AI development and deployment. For instance, Doe et al. [20] investigated the ethical implications of biased outputs generated by language models, emphasizing the need for algorithmic transparency and bias mitigation strategies. Similarly, Smith and Jones [21] explored the ethical deliberations related to employing generative AI in content generation, advocating for responsible AI methodologies to guarantee equity and responsibility.

2.2 Societal Implications

The widespread dissemination of synthetic media, such as deepfakes, created by generative AI systems presents significant societal ramifications. Deepfakes, portraying individuals or occurrences with remarkable accuracy, possess the capability to mislead and manipulate audiences, diminish confidence in media and establishments, and disrupt democratic procedures. The extensive proliferation of deepfakes heightens the difficulty of countering misinformation in the digital realm, emphasizing the immediate requirement for effective detection and mitigation strategies. Garcia and Martinez [22] investigated the psychological effects of exposure to deepfake content, highlighting the potential risks to mental health and well-being.

2.3 Security Risks

Generative AI systems introduce security risks, particularly concerning cybersecurity and privacy. The creation of synthetic images or videos that are virtually identical to genuine ones can be utilized for nefarious motives, including identity theft, extortion, or espionage. Additionally, employing generative AI for data generation in sensitive areas raises apprehensions about the potential abuse or unauthorized retrieval of synthetic data [23]. Protecting against these security risks requires robust cybersecurity measures and regulatory safeguards.

Research endeavors have aimed to tackle the security ramifications of generative AI systems. Tan et al. [23] conducted a comprehensive analysis of security vulnerabilities in generative AI models, proposing defense mechanisms to mitigate the risk of adversarial attacks. Additionally, Chen and Wang [24] examined the privacy implications of generative AI-generated data, advocating for encryption and differential privacy techniques to safeguard sensitive information. Table 1 elaborates catagorical AI Risks.

Table 1. Comparative Table of Generative AI Risks

Risk Category	Description	References
Ethical Concerns	Biased or misleading content generated by AI systems	[20-21] [25]
	Propagation of stereotypes and	[22] [4] [23]

	misinformation in generated content	
Societal Implications	Impact of synthetic media (e.g., deepfakes) on trust in media and institutions	[18] [24] [20]
	Potential for deception and manipulation of individuals through synthetic media	[21] [22] [25]
Security Risks	The utilization of synthetic media for malevolent intentions such as identity theft or espionage.	[4] [23] [7]
	Security implications of generating synthetic data in sensitive domains (e.g., healthcare, finance)	[24] [20] [21]

3. Challenges in Risk Assessment

Assessing risks associated with generative AI systems presents several challenges, primarily due to their complex and dynamic nature. Traditional risk assessment frameworks may not be suitable for capturing the nuanced and evolving risks posed by generative AI. Some of the key challenges include:

3.1 Unpredictable Outputs

Generative AI systems have the capability to produce outputs that are often unpredictable and may deviate from expected norms. This unpredictability poses significant challenges in risk assessment, as it becomes difficult to anticipate and mitigate potential adverse outcomes. As Ali et al. [26] highlight in their seminal work on deep learning, generative models introduce an element of randomness and creativity, leading to uncertainties about the nature and implications of generated outputs.

Moreover, the lack of control over the specific characteristics of generated outputs can pose ethical and regulatory challenges. For example, within the domain of natural language generation, generative AI models might unintentionally generate text containing prejudiced or unsuitable language, potentially resulting in harm to the reputation of organizations utilizing such systems [4]. The unpredictable nature of generative AI outputs underscores the importance of robust risk assessment frameworks that can effectively identify and mitigate potential risks.

3.2 Lack of Interpretability

A notable hurdle in evaluating risks linked to generative AI systems is the absence of interpretability in numerous AI models. Frequently, generative AI approaches function akin to "black boxes," where the internal workings and decision-making procedures are obscure and challenging to decipher. This absence of clarity impedes efforts to assess risks by concealing the elements influencing model behavior and the reasoning behind generated outcomes.

3.3 Rapid Technological Advancements

The field of generative AI is characterized by rapid technological advancements and continuous innovation, driven by ongoing research and development efforts. New techniques, architectures, and models are constantly being introduced, leading to frequent updates and improvements in generative AI capabilities. While these advancements hold promise for unlocking new applications and capabilities, they also present challenges for risk assessment practitioners.

Zellers et al. [27] elaborated the progress of AI research and the it’s hinderances for risk evaluation techniques. The constant developments in generative AI demand risk assessment frameworks to adjust and develop accordingly, underscoring the need for continual research and advancement in the domain of AI risk management. Failure to keep pace with these advancements could result in outdated risk assessment practices that fail to adequately capture emerging risks associated with novel techniques and methodologies.

4. Risk Assessment Frameworks

In response to the challenges posed by generative AI systems, researchers and practitioners have developed various frameworks aimed at assessing and mitigating associated risks. These frameworks provide structured approaches for identifying, analyzing, and mitigating risks, considering the

unique characteristics of generative AI. In this section, we explore some notable frameworks and their contributions to risk assessment in generative AI systems.

4.1 Ethical AI Framework

The Ethical AI Framework stands out as a notable approach for evaluating risks in generative AI systems. This framework underscores the ethical aspects associated with AI creation and application, focussing on principles like equity, openness, and responsibility. As emphasized by Jobin and colleagues [28], ethical deliberations are crucial in the development and execution of AI systems, especially those possessing generative capacities.

The Ethical AI Framework offers guidance and principles for assessing the ethical ramifications of generative AI technologies and ensuring conformity with societal values and standards. By integrating principles like equity and openness into the creation and implementation of AI systems, entities can alleviate risks linked to prejudice, differentiation, and unforeseen outcomes. The framework underscores the significance of involving stakeholders and fostering interdisciplinary cooperation to tackle ethical dilemmas and advance accountable AI advancement.

4.2 Adversarial Testing Framework:

Another important framework for assessing risks in generative AI systems is the Adversarial Testing Framework. This framework involves subjecting generative AI systems to adversarial testing scenarios to identify vulnerabilities and enhance model robustness against malicious attacks. Adversarial testing techniques, such as adversarial examples and adversarial training, aim to expose weaknesses in AI systems and improve their resilience to adversarial manipulation.

Brundage et al. [29] discuss the significance of adversarial testing in uncovering vulnerabilities in generative AI systems and mitigating security risks. By systematically testing AI systems against adversarial attacks, organizations can identify potential weaknesses and implement appropriate countermeasures to enhance model robustness and security. The Adversarial Testing Framework complements traditional risk assessment methodologies by focusing on proactive measures to defend against emerging threats and vulnerabilities.

4.3 Human-in-the-Loop Framework:

This Framework offers an alternative method for evaluating risks in generative AI systems. This framework incorporates human supervision and feedback loops into the generative process, allowing for continuous monitoring and modification of model behavior in response to ethical and societal factors. Mittelstadt et al. [30] stressed the significance of human engagement in AI systems to guarantee compatibility with human values and preferences.

Through the integration of human supervision and feedback mechanisms, this framework improves transparency, accountability, and manageability in generative AI systems. Human-in-the-loop approaches enable stakeholders to intervene in the generative process when necessary, addressing ethical concerns and mitigating potential risks associated with autonomous decision-making. This framework promotes collaborative decision-making between AI systems and human operators, fostering trust and confidence in AI-driven processes.

5. Comprehensive study on Managing AI Risks in Finance Sector

In the past few years, the finance sector has experienced a swift uptake of artificial intelligence (AI) technologies to streamline operations, elevate decision-making procedures, and enhance customer satisfaction. However, alongside the benefits, the integration of AI in financial services introduces complex risks that must be carefully managed to safeguard against potential harm to individuals, organizations, and the broader financial system. This study explores how a global consulting firm specializing in risk management and compliance, assists financial institutions in identifying, assessing, and mitigating AI-related risks.

5.1 Significance of the study

As AI technologies become more intertwined with financial operations, the likelihood of negative consequences, such as biased decision-making, breaches of data, and failure to comply with regulations, increases. Financial institutions can utilize their proficiency in risk management to

navigate the intricate realm of AI risks and ensure that AI technologies are used responsibly and ethically.

5.2 Sample Size and Questionnaire Analysis

A comprehensive risk assessment study was conducted by administering questionnaires to 50 key stakeholders within the financial institution. The sample size of 50 respondents encompassed a diverse cross-section of individuals directly involved in AI implementation and oversight within the organization, including executives, risk managers, data scientists, and compliance officers.

5.2.1 Questionnaire Content:

The survey included a range of questions, both open-ended and closed-ended, crafted to gather perspectives on different aspects of AI risk. These aspects encompassed issues such as bias and fairness, security vulnerabilities, compliance with regulations, and the effectiveness of current risk management strategies. All questionnaires used in this study are presented in Annexure A. Sample questions are as follows:

1. Are you informed about the possible hazards linked to implementing generative AI models in credit scoring procedures?
2. How do you ensure the fairness and transparency of AI-driven credit decisions?
3. What actions have been taken to safeguard sensitive customer data from unauthorized access or manipulation?

5.3 Survey Results:

The analysis of questionnaire responses yielded several key findings and insights into AI-related risks within the financial institution:

- **Bias and Fairness Concerns:** The survey findings indicated a broad recognition of the possibility for AI algorithms to sustain biases inherent in past data, especially in credit scoring and lending procedures. Stakeholders voiced worries regarding the fairness and clarity of decisions driven by AI, emphasizing the necessity for thorough testing, validation, and continual monitoring to reduce bias and guarantee fair results.
- **Security Risks:** Security emerged as a paramount concern, with respondents highlighting the vulnerability of AI systems to cyber threats, data breaches, and adversarial attacks. Incorporating AI into vital financial operations, like identifying fraud and assessing risks, heightens the potential consequences of security breaches. Therefore, it is essential to implement strong safeguards, encryption protocols, and intrusion detection systems to protect sensitive data and uphold confidence in financial systems.
- **Regulatory Compliance:** Compliance with regulatory requirements emerged as a significant challenge, particularly in highly regulated jurisdictions where stringent data protection and consumer privacy laws govern AI deployment. Respondents cited difficulties in navigating complex regulatory landscapes, interpreting regulatory guidance, and ensuring alignment with evolving best practices, highlighting the importance of actively involving regulatory authorities and industry stakeholders to ensure compliance and reduce legal risks.
- **Risk Management Practices:** The questionnaire analysis revealed variations in risk management maturity across different functional areas, certain departments exhibit a deeper comprehension of AI risks and take proactive steps to address them effectively. However, gaps in risk governance, accountability, and oversight were identified, highlighting the importance of establishing clear policies, procedures, and escalation protocols to address emerging risks and ensure organizational resilience in the face of AI-related challenges.

5.4 Implications for Risk Management:

The results of the study emphasize the diverse aspects of AI risks within the finance sector and highlight the urgent need for financial institutions to embrace a comprehensive and proactive strategy for risk mitigation.

By leveraging expertise in risk assessment, regulatory compliance, and technology governance, financial organizations can strengthen their resilience to AI risks and enhance trust. Key implications for risk management include:

- Giving precedence to equity, openness, and responsibility in AI development and utilization.
- Putting in place strong security protocols to safeguard against cyber threats and breaches of data.
- Collaborating with regulatory bodies and industry colleagues to remain updated on evolving regulatory standards and optimal methodologies.

- Strengthening risk management frameworks, procedures, and capacities to accurately recognize, evaluate, and alleviate AI-associated risks.

6. Conclusions

This study presents the critical aspects of AI and associated risks, offering insights into the challenges and implications across ethical, societal, and security domains. From this comprehensive examination, the following conclusions can be drawn

- Giving priority to equity, openness, and responsibility during the creation and utilization of AI systems is crucial for addressing ethical worries linked with biased content generation and the spread of stereotypes and false information.
- The implementation of strong cybersecurity measures, like encryption protocols and intrusion detection systems, plays a pivotal role in defending against security threats such as data breaches, identity theft, and espionage.
- Actively engaging with regulatory bodies and industry counterparts is essential to ensure adherence to evolving legal and regulatory standards, particularly in heavily regulated sectors like finance.
- Strengthening governance frameworks, procedures, and capabilities is essential for effectively recognizing, evaluating, and mitigating risks associated with AI, thereby fostering organizational resilience and bolstering confidence in AI-driven decision-making processes.
- Collaboration among various stakeholders, including AI developers, ethicists, policymakers, and regulators, is crucial for addressing the multifaceted challenges posed by generative AI and promoting responsible development and deployment of AI.
- The comprehensive examination undertaken in the finance sector highlights the significance of equity, transparency, and security in AI deployment, in addition to regulatory compliance and effective risk management, as critical aspects for financial institutions navigating the complexities of AI risks.
- Ongoing research and advancement in AI risk management are imperative to keep abreast of the rapid advancements in generative AI technologies and effectively tackle emerging challenges.

References

[1] Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. Social media + Society, 6(1).

[2] Lv, Z. (2023). Generative artificial intelligence in the metaverse era. Cogn Robot, 3, 208-217.

[3] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Aaron. C., & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

[4] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1(8), 9.

[5] Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2021). Green AI. arXiv preprint arXiv:2007.01859.

[6] Liang, Y., Zhu, S., Zhang, C., & Hu, B. (2021). Legal Implications of Generative AI-Generated Content: Challenges and Opportunities. Journal of Law and Technology, 2(1), 45-58.

[7] Wang, Z., & Zhang, Z. (2020). Security Implications of Generative AI in Sensitive Domains: Challenges and Countermeasures. Journal of Cybersecurity, 1(1), 34-47.

[8] Al-khazraji, S., Saleh, H., Khalid, A., & Mishkhal, I. (2023). Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications. Eurasia Proc Sci Tech Eng Math, 23, 429-441.

[9] Oksanen, A., Cvetkovic, A., Akin, N., Latikka, R., Bergdahl, J., Chen, Y., & Savela, N. (2023). Artificial intelligence in fine arts: A systematic review of empirical research. Comput Human Behav: Artif Humans, 1(2), 100004.

[10] Smith, A., Johnson, B., & Williams, C. (2022). Generative Artificial Intelligence and Financial Fraud Detection. J Financ Technol, 12(3), 45-58.

[11] Jones, D., & Lee, E. (2023). Market Risks in AI-Driven Trading Algorithms. J Econ Finance, 15(2), 110-125.

[12] Chen, F., & Wang, G. (2023). Ethical Considerations in Credit Scoring Using Generative AI. J Bank Financ, 25(4), 220-235.

[13] Gupta, R., Sharma, S., & Patel, M. (2023). Security Vulnerabilities of AI Chatbots in Customer Service. J Comput Secur, 8(1), 75-89.
[14] Kim, H., Park, J., & Lee, S. (2023). Societal Implications of Generative AI in Personal Finance Management. J Econ Soc, 20(3), 315-330.
[15] Li, Y., & Zhang, L. (2023). Regulatory Challenges in Compliance Monitoring with Generative AI. J Regul Econ, 18(2), 150-165.
[16] Wu, Q., & Liu, K. (2023). Legal Liabilities of AI-Driven Investment Recommendations. J Law Finance, 30(1), 40-55.
[17] Park, M., & Choi, S. (2023). Impact of Generative AI on Financial Stability. J Financ Stability, 5(4), 280-295.
[18] Wang, & Huang. (2023). Role of Generative AI in Asset Pricing Models. J Asset Pricing, 12(1), 60-75.
[19] Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations, and future scope. Internet Things Cyber-Phys Syst, 3, 121-154.
[20] Doe, J., Smith, A., & Jones, B. (2020). Ethical Considerations in Generative AI: A Framework for Responsible Development and Deployment. Journal of AI Ethics, 1-15.
[21] Smith, A., & Jones, B. (2021). Ethical Considerations in AI Content Creation: Challenges and Recommendations. Journal of AI Ethics, 2(1), 56-72.
[22] Garcia, C., & Martinez, D. (2021). Psychological Effects of Exposure to Deepfake Content: A Study on Perception and Trust. Journal of Digital Media Studies, 5(2), 78-92.
[23] Tan, Z., Zhang, X., & Li, Y. (2020). Privacy-Preserving Generative Adversarial Networks for Synthetic Data Generation: A Comprehensive Review. Journal of Privacy and Security, 3(2), 123-137.
[24] Chen, X., & Wang, X. (2021). Privacy-Preserving Generative Adversarial Networks for Data Augmentation: A Survey. arXiv preprint arXiv:2101.01062.
[25] Johnson, R., Garcia, C., & Martinez, D. (2020). Understanding the Societal Implications of Deepfakes: A Comprehensive Analysis. Journal of Media Ethics, 4(3), 112-128.
[26] Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. Inf Fusion, 99, 101805.
[27] Zellers, R., Bisk, Y., Schwartz, R., & Choi, Y. (2019). HellaSwag: Can a Machine Really Finish Your Sentence?. arXiv preprint arXiv:1905.07830.
[28] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.
[29] Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Zeitzoff, T. (2020). Toward trustworthy AI development: Mechanisms for supporting verifiable claims. arXiv preprint arXiv:2004.07213.
[30] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2019). The ethics of algorithms: Mapping the debate. Big Data & Society, 6(2), 2053951716679679.

Annexure A: Questionnaire

1. Are you aware of the potential risks associated with deploying generative AI models in credit scoring processes?

a. Yes
b. No
c. Not sure

2. How do you ensure the fairness and transparency of AI-driven credit decisions?

a. Regular monitoring and auditing of model outputs
b. Incorporation of fairness metrics into model development
c. Collaboration with domain experts to identify and mitigate biases
d. Other (please specify)

3. What measures are in place to protect sensitive customer data from unauthorized access or manipulation?

a. Encryption of data at rest and in transit
b. Access controls and authentication mechanisms
c. Regular security audits and vulnerability assessments
d. Compliance with data protection regulations (e.g., GDPR, CCPA)

e. Other (please specify)

4. Have you encountered any security incidents or data breaches related to AI systems in the past?
a. Yes
b. No
c. Not applicable

5. How do you ensure compliance with regulatory requirements governing the use of AI in financial decision-making?
a. Regular review of regulatory guidance and updates
b. Engagement with legal and compliance teams to interpret and implement regulations
c. Integration of regulatory compliance checks into AI development lifecycle
d. Other (please specify)

6. Are there any specific challenges or concerns you have regarding the use of AI in financial decision-making processes?
a. Bias and fairness concerns
b. Security vulnerabilities
c. Regulatory compliance issues
d. Ethical considerations
e. Other (please specify)

7. What measures do you believe should be implemented to improve the management of AI-related risks within the organization?
a. Enhanced training and awareness programs for employees
b. Establishment of clear policies and procedures for AI governance
c. Investment in advanced security technologies and protocols
d. Collaboration with external experts and industry peers
e. Other (please specify)

8. How confident are you in the organization's ability to effectively manage AI-related risks?
a. Very confident
b. Somewhat confident
c. Neutral
d. Not very confident
e. Not confident at all

9. Are there any additional comments or suggestions you would like to provide regarding AI risk management within the organization?