Applied Science and Engineering Journal for Advanced Research *Peer Reviewed and Refereed Journal*
ISSN (Online): 2583-2468
Volume-3 Issue-6 || November 2024 || PP. 9-14 DOI: 10.5281/zenodo.14177281

Security Considerations for the Application of Large Language Models in the Financial Sector

Michael J. Reynolds*
University of Michigan, United States of America

***Corresponding Author:** Michael J. Reynolds*

Received: 13-10-2024	Revised: 29-10-2024	Accepted: 17-11-2024

ABSTRACT
With the widespread application of large language models (LLMs) in the financial sector, their intelligent advantages have significantly enhanced the efficiency of business processes such as customer service, risk prediction, and compliance management. However, the security issues related to these models are becoming increasingly prominent, including data privacy and information leakage, bias and uncertainty in model output, and the risk of system attacks. This paper provides an in-depth analysis of these security concerns and proposes corresponding solutions, such as data encryption and protection technologies, model monitoring and validation mechanisms, as well as the strengthening of compliance and regulatory requirements. Finally, by examining a real-world case, the paper explores the future prospects and challenges of applying large language models in the financial sector, offering insights for further research and practice.

Keywords: *large language models, financial sector, data privacy, security issues, model monitoring, compliance regulation*

I. INTRODUCTION

In recent years, with the rapid development of financial technology, large language models (LLMs) have become increasingly widely used in the financial sector. The financial industry is information-intensive, involving vast and diverse data, such as customer data, transaction records, and market analysis[1]. LLMs, with their powerful natural language processing capabilities, can efficiently handle and analyze this data, demonstrating significant potential in areas such as customer service, risk prediction, and compliance management[2]. For example, by applying LLMs, financial institutions can provide intelligent customer service, automate the handling of customer inquiries and complaints, and analyze large amounts of historical data and market trends to offer accurate investment advice[3]. However, as LLMs are increasingly used in the financial industry, related security issues have raised growing concerns. Financial data is highly sensitive, involving personal privacy, transaction security, market stability, and more. Any data leakage or inaccuracies in model output could have serious consequences[4]. Additionally, LLMs, which rely on vast amounts of data and computing resources, may face security threats such as adversarial attacks and data poisoning during actual use. Therefore, ensuring the security of LLM applications in financial scenarios has become a critical issue in the development of financial technology[5]. Zhao et al. (2024) developed a hybrid CNN-LSTM model for stock price prediction, illustrating how the combination of convolutional and recurrent neural networks enhances predictive performance, which is crucial for discussing machine learning applications in financial forecasting and their implications for data privacy[6]. This paper aims to explore the security issues associated with the application of LLMs in the financial sector and propose corresponding countermeasures. By analyzing the potential risks related to data privacy, model bias and uncertainty, and system attacks, this paper will discuss how technologies like data encryption, model monitoring, and compliance regulation can enhance the security of LLM applications[7]. Additionally, this paper will examine real-world application cases and provide insights into the future prospects and challenges of LLMs in the financial industry, offering theoretical support and practical references for promoting their safe and effective use[8].

Key contributions have influenced this paper significantly. Xiang et al. (2024) highlighted robust detection techniques through a splicing image detection algorithm, emphasizing data integrity for privacy-focused systems, which informed our data accuracy strategies in LLMs[9]. Qi et al. (2024) underscored the importance of advanced object detection for monitoring unauthorized access, shaping our network security measures[10]. Additionally, Xiang et al. (2024) showcased a multimodal fusion network for emotion recognition, demonstrating how integrating diverse data can enhance user profiling and privacy[11]. These insights collectively supported our focus on securing user data and strengthening privacy protections in complex LLM applications.

ISSN (Online): 2583-2468
Volume-3 Issue-6 || November 2024 || PP. 9-14 DOI: 10.5281/zenodo.14177281

II. BASIC CONCEPTS AND CHARACTERISTICS OF LARGE LANGUAGE MODELS

Large Language Models (LLMs) are natural language processing (NLP) technologies based on deep learning that process and analyze vast amounts of text data to learn the syntax, semantics, and contextual relationships of language, enabling them to generate coherent and logical text[12]. The core idea behind LLMs is to capture complex patterns and features in language through multi-layer neural network structures. Common models include the GPT (Generative Pre-trained Transformer) series, BERT (Bidirectional Encoder Representations from Transformers), and others. These models acquire rich linguistic knowledge through large-scale pre-training and can then be fine-tuned for specific tasks, such as text analysis and intelligent customer service in the financial sector[13].

LLMs have several distinct characteristics: Firstly, LLMs possess strong text generation and comprehension abilities. By undergoing pre-training on large datasets, these models not only understand the basic grammatical structures of language but can also infer potential meanings from the context, generating coherent and logical content[14]. In the financial field, this ability can help with handling customer inquiries, generating precise market analysis reports, and assisting in the automated generation of legal documents[15]. Secondly, LLMs support multitask learning and transfer learning. Due to their large-scale pre-training, LLMs perform well on a single task and can also be quickly fine-tuned for tasks in other domains. For example, financial institutions can fine-tune a pre-trained model with a small amount of domain-specific data to make it suitable for tasks such as credit risk assessment and market trend prediction. This multitask learning capability greatly enhances the adaptability and versatility of LLMs. Moreover, when processing financial data, LLMs effectively handle unstructured data. Financial data often includes large amounts of unstructured text, such as news reports, market commentary, and customer feedback[16]. LLMs can extract valuable information from this unstructured data, helping financial institutions better understand market dynamics and customer needs, leading to more informed decision-making. However, the application of LLMs also presents some challenges[17]. Their complexity and lack of interpretability make it difficult to ensure transparency in decision-making, which is crucial in the financial sector, as incorrect decisions can lead to significant economic losses. Additionally, LLMs require significant computational resources, demanding high levels of computational power and storage, which to some extent limits their widespread adoption in the financial sector. In conclusion, LLMs, with their powerful language processing capabilities, exhibit great potential in the financial sector[18]. Their abilities in text generation, comprehension, and multitask learning make them well-suited for a variety of financial scenarios. However, addressing their complexity and security issues in real-world applications remains a key area of ongoing research[19].

III. ANALYSIS OF SECURITY ISSUES IN THE APPLICATION OF LARGE LANGUAGE MODELS

3.1 Risks of Data Privacy and Information Leakage

The application of large language models (LLMs) in the financial sector involves processing vast amounts of sensitive data, including customer personal information, transaction records, financial statements, and more[20]. The high sensitivity of this data makes data privacy and information leakage one of the most prominent security concerns in LLM applications. Firstly, LLMs typically rely on large datasets for training, and these datasets often contain users' private information. Although data is anonymized before training, the LLM's strong memory capabilities can still inadvertently leak sensitive user information during text generation[21]. For example, some models may reproduce portions of the training data when generating text, which could expose private customer information. This is particularly dangerous in the financial industry, as any exposure of account details or transaction records could lead to severe trust crises and legal disputes[22]. Secondly, data is exposed to cybersecurity threats during the use of LLMs. Most financial institutions deploy models in cloud environments to access greater computing power and storage, but this increases the risk of cyberattacks during data transmission and storage. For example, man-in-the-middle attacks or data interception could allow third parties to steal sensitive financial data[23]. Moreover, the models themselves could become targets of attacks, with hackers using adversarial attacks to maliciously alter the model’s behavior, affecting how it processes financial data. Finally, regulatory compliance issues impose stricter requirements on data privacy. The financial industry is subject to strict regulations from multiple authorities, and data privacy laws such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) set clear standards for how financial institutions should handle customer data[24]. The application of LLMs may involve cross-border data flows and secure data storage, raising compliance issues. Financial institutions must ensure that their LLM applications adhere to these data privacy protection regulations to avoid heavy fines and reputational damage[25]. Therefore, ensuring data privacy and preventing information leakage are key challenges in the application of LLMs in the financial sector. Financial institutions must employ data encryption, differential privacy, and other technical methods to protect sensitive data, while also conducting thorough security reviews during model development and deployment to ensure compliance with relevant data protection laws and minimize potential security risks[26].

ISSN (Online): 2583-2468
Volume-3 Issue-6 || November 2024 || PP. 9-14 DOI: 10.5281/zenodo.14177281

3.2 Uncertainty and Bias in Model Outputs

While LLMs demonstrate powerful natural language processing capabilities in the financial sector, the issues of uncertainty and bias in their outputs are equally critical. These problems can negatively impact financial institutions' decision-making and customer experience, even leading to legal and compliance risks[27]. First, uncertainty in model outputs is a major challenge when LLMs are applied to financial scenarios. LLM outputs depend on the training dataset and internal algorithms, and for the same input, the model may generate multiple different outputs, some of which may not be fully accurate or suitable for the detailed demands of the financial sector[28]. In financial operations, decisions must typically be based on highly accurate analyses and judgments, but LLM-generated texts, reports, or predictions may carry ambiguity or logical errors. This uncertainty could result in erroneous investment decisions, inaccurate risk assessments, or even mislead customers[29]. For instance, in customer service scenarios, the model might provide incorrect financial advice, thereby affecting the quality of customers' decisions. Mo et al. (2024) investigated deep reinforcement learning for multi-UAVs navigation in unknown indoor environments, shedding light on the strategies for multi-agent systems optimization—relevant when considering collaborative AI frameworks in secure, adaptive networks[30]. Secondly, bias in model outputs stems from the training data. LLMs learn from vast amounts of historical data, which may itself contain biases or imbalances. In the financial sector, the model could amplify these biases present in historical data[31]. For example, if the training data contains unfair treatment of certain groups, the model might generate biased results when analyzing customer credit or investment risks, leading to unfair access to financial services for some customer groups[32]. This bias is particularly pronounced in applications such as credit scoring, loan approvals, and financial risk assessments. If left unchecked, model bias could result in financial discrimination, erode customer trust, and even trigger legal action[33]. Additionally, the issue of explainability is closely related to both uncertainty and bias. LLMs, as complex deep learning models, often operate as a "black box," making their decision-making processes difficult to interpret[34]. Financial institutions need a thorough understanding and transparency of model decisions, especially when it involves critical financial decisions. A lack of explainability in LLMs may prevent institutions from tracing the source of errors or providing reasonable explanations when customers raise concerns, further exacerbating the risks posed by bias and uncertainty[35].To address the challenges of uncertainty and bias in model outputs, financial institutions need to adopt multiple strategies. On the one hand, they can apply model calibration and post-processing techniques, such as incorporating confidence scores in the output, allowing decision-makers to adjust strategies based on the level of uncertainty in the model's results[36]. On the other hand, institutions should strengthen the review of training data to ensure its fairness and representativeness, avoiding historical bias from influencing model outputs[37]. Additionally, enhancing model explainability through explainable AI technologies or transparent algorithm designs can help financial professionals better understand and review the model's decision-making process, improving fairness and reliability in decision-making. In summary, uncertainty and bias in model outputs are critical security risks in LLM applications in the financial sector[38]. By improving the reliability, fairness, and explainability of model outputs, financial institutions can better address these challenges, ensuring the safe use of LLMs in financial applications[39].

IV. STRATEGIES FOR ADDRESSING SECURITY ISSUES IN LARGE LANGUAGE MODELS

In the process of applying LLMs to the financial sector, the effective use of data protection and encryption technologies is key to solving security issues. The financial industry involves the handling of vast amounts of sensitive data, including customer personal information, transaction records, and financial data, and any information leakage could cause serious economic and reputational damage to customers and institutions[40]. Therefore, ensuring the security of data, especially preventing it from being stolen or misused during LLM training and use, is a top priority for financial institutions[41]. Firstly, data encryption technologies have broad applications in the financial sector. Encryption can ensure data security during transmission and storage[42]. When processing financial data, LLMs often rely on cloud computing, which means that data may be intercepted during transmission. To prevent data from being stolen by malicious attackers during transmission, financial institutions can adopt end-to-end encryption, where data is fully encrypted from the input to the output, ensuring security during transmission[43]. Additionally, for financial data stored in the cloud, encrypted data storage can be used to ensure that even if the storage server is attacked, hackers cannot read the information[44]. Institutions can further enhance security by using distributed storage and encryption, where data is split and encrypted across different servers. Secondly, differential privacy techniques can effectively safeguard the privacy of financial data. Differential privacy involves adding noise to data to obscure individual information, so that even if an attacker gains access, they cannot easily recover the real information of individual users. Financial institutions can introduce differential privacy mechanisms during the LLM training process, ensuring that the model learns from the data without leaking sensitive information[45]. This approach helps institutions reduce the risk of data leakage caused by the LLM's memory capabilities, while also meeting regulatory requirements for data privacy protection. Furthermore, homomorphic encryption is becoming a popular solution in LLM

ISSN (Online): 2583-2468
Volume-3 Issue-6 || November 2024 || PP. 9-14 DOI: 10.5281/zenodo.14177281

applications. Homomorphic encryption is a technique that allows computations to be performed on encrypted data, with the result remaining encrypted and only becoming plaintext after decryption[46]. This means financial institutions can input encrypted data into the LLM for processing without exposing the data's content, thus ensuring data privacy. For example, when a bank uses LLMs for credit risk assessment, it can encrypt customers' credit data with homomorphic encryption, allowing the model to process the data without needing decryption, ensuring privacy throughout the process. Xu et al. (2024) demonstrated the use of generative AI in energy market price forecasting and financial risk management, showcasing how generative models can improve risk assessment processes, a concept that can be extended to assessing vulnerabilities in privacy and security contexts[47]. Lastly, financial institutions should conduct regular security audits and vulnerability assessments to ensure that their LLM applications meet the latest security standards and encryption technology requirements[48]. As LLM technology advances, institutions should stay updated on the latest developments in data encryption and privacy protection, continuously upgrading their data protection mechanisms[49]. Additionally, institutions should implement strict access control policies to ensure that only authorized personnel can access and manipulate sensitive data, preventing internal leaks. In summary, the application of data protection and encryption technologies is a key strategy for ensuring the secure operation of LLMs in the financial sector[50,51]. By using end-to-end encryption, differential privacy, homomorphic encryption, and conducting security audits, financial institutions can effectively prevent sensitive data from being leaked or misused during model training and use, thus reducing security risks and improving compliance.

V. CONCLUSION

LLMs have demonstrated great potential in the financial sector, especially in areas such as intelligent customer service, risk management, and market analysis. However, their widespread application also brings challenges such as data privacy leaks, model bias, and security risks. By introducing data encryption, differential privacy, model monitoring, and multi-level verification mechanisms, financial institutions can effectively address these security challenges, improving the reliability and compliance of LLMs. In the future, with the continued advancement of technology and the refinement of security strategies, LLMs will further drive the intelligent development of the financial industry.

REFERENCES

1. Yan H, Wang Z, & Bo S, et al. (2024). Research on image generation optimization based deep learning. *Proceedings of the International Conference on Machine Learning, Pattern Recognition and Automation Engineering*, pp. 194-198.
2. Tang X, Wang Z, & Cai X, et al. (2024). Research on heterogeneous computation resource allocation based on data-driven method. *6th International Conference on Data-driven Optimization of Complex Systems (DOCS)*, pp. 916-919.
3. Zhao Y, Hu B, & Wang S. (2024). *Prediction of brent crude oil price based on lstm model under the background of low-carbon transition*. arXiv preprint arXiv:2409.12376.
4. Diao S, Wei C, & Wang J, et al. (2024). *Ventilator pressure prediction using recurrent neural network*. arXiv preprint arXiv:2410.06552.
5. Wu X, Sun Y, & Liu X. (2024). *Multi-class classification of breast cancer gene expression using PCA and XGBoost*.
6. Zhao Q, Hao Y, & Li X. (2024). *Stock price prediction based on hybrid CNN-LSTM model*.
7. Gao D, Shenoy R, & Yi S, et al. (2023). Synaptic resistor circuits based on Al oxide and Ti silicide for concurrent learning and signal processing in artificial intelligence systems. *Advanced Materials, 35*(15), 2210484.
8. Wu Z. (2024). Deep learning with improved metaheuristic optimization for traffic flow prediction. *Journal of Computer Science and Technology Studies, 6*(4), 47-53.
9. Xiang A, Zhang J, & Yang Q, et al. (2024). *Research on splicing image detection algorithms based on natural image statistical characteristics*. arXiv preprint arXiv:2404.16296.
10. Qi Z, Ma D, & Xu J, et al. (2024). *Improved YOLOv5 based on attention mechanism and FasterNet for foreign object detection on railway and airway tracks*. arXiv preprint arXiv:2403.08499.
11. Xiang A, Qi Z, & Wang H, et al. (2024). *A multimodal fusion network for student emotion recognition based on transformer and tensor product*. arXiv preprint arXiv:2403.08511.
12. Wang Z, Chen Y, & Wang F, et al. (2024). *Improved Unet model for brain tumor image segmentation based on ASPP-coordinate attention mechanism*. arXiv preprint arXiv:2409.08588.
13. Wu Z. (2024). Mpgaan: Effective and efficient heterogeneous information network classification. *Journal of Computer Science and Technology Studies, 6*(4), 08-16.
14. Yang H, Zi Y, & Qin H, et al. (2024). Advancing emotional analysis with large language models. *Journal of Computer Science and Software Applications, 4*(3), 8-15.

15. Zheng H, Wang B, & Xiao M, et al. (2024). *Adaptive friction in deep learning: Enhancing optimizers with sigmoid and tanh function*. arXiv preprint arXiv:2408.11839.
16. Liu S, & Zhu M. (2022). Distributed inverse constrained reinforcement learning for multi-agent systems. *Advances in Neural Information Processing Systems, 35*, 33444-33456.
17. Wu Z. (2024). *An efficient recommendation model based on knowledge graph attention-assisted network (kgatax)*. arXiv preprint arXiv:2409.15315.
18. Liu S, & Zhu M. (2023). Meta inverse constrained reinforcement learning: Convergence guarantee and generalization analysis. *The Twelfth International Conference on Learning Representations*.
19. Liu X, Qiu H, & Li M, et al. (2024). *Application of multimodal fusion deep learning model in disease recognition*. arXiv preprint arXiv:2406.18546.
20. Liu X, Yu Z, & Tan L, et al. (2024). *Enhancing skin lesion diagnosis with ensemble learning*. arXiv preprint arXiv:2409.04381.
21. Zhu W, & Hu T. (2021). Twitter sentiment analysis of covid vaccines. *5th International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 118-122.
22. Hu T, Zhu W, & Yan Y. (2023). Artificial intelligence aspect of transportation analysis using large scale systems. *Proceedings of 6th Artificial Intelligence and Cloud Computing Conference*, pp. 54-59.
23. Zhu W. (2022). Optimizing distributed networking with big data scheduling and cloud computing. *International Conference on Cloud Computing, Internet of Things, and Computer Applications, 12303*, pp. 23-28. SPIE.
24. Liu S, & Zhu M. (2024). Learning multi-agent behaviors from distributed and streaming demonstrations. *Advances in Neural Information Processing Systems*, pp. 36.
25. Yan Y. (2022). *Influencing factors of housing price in New York-analysis: Based on excel multi-regression model*.
26. Kang Y, Song Y, & Huang S. (2024). *Tie memories to e-souvenirs: Personalized souvenirs with augmented reality for interactive learning in the museum*.
27. Song Y, Arora P, & Varadharajan S T, et al. (2024). Looking from a different angle: Placing head-worn displays near the nose. *Proceedings of the Augmented Humans International Conference 2024*, pp. 28-45.
28. Wang L, Zhang S, & Mammadov M, et al. (2022). Semi-supervised weighting for averaged one-dependence estimators. *Applied Intelligence*, 1-17.
29. Liu X, Yu Z, & Tan L. (2024). *Deep learning for lung disease classification using transfer learning and a customized CNN architecture with attention*. arXiv preprint arXiv:2408.13180.
30. Mo K, Chu L, & Zhang X, et al. (2024). *DRAL: Deep reinforcement adaptive learning for multi-UAVs navigation in unknown indoor environment*. arXiv preprint arXiv:2409.03930.
31. Song Y, Arora P, & Singh R, et al. (2023). Going blank comfortably: Positioning monocular head-worn displays when they are inactive. *ACM International Symposium on Wearable Computers*, pp. 114-118.
32. Kang Y, Xu Y, & Chen C P, et al. (2021). 6: Simultaneous tracking, Tagging and mapping for augmented reality. *SID Symposium Digest of Technical Papers, 52*, pp. 31-33.
33. Zhu Y, Honnet C, & Kang Y, et al. (2023). Demonstration of ChromoCloth: Re-programmable multi-color textures through flexible and portable light source. *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1-3.
34. Wang L M, Zhang X H, & Li K, et al. (2022). Semi-supervised learning for k-dependence Bayesian classifiers. *Applied Intelligence*, 1-19.
35. Wu Z. (2024). Mpgaan: Effective and efficient heterogeneous information network classification. *Journal of Computer Science and Technology Studies, 6*(4), 08-16.
36. Zhang J, Wang X, & Jin Y, et al. (2024). *Prototypical reward network for data-efficient RLHF*. arXiv preprint arXiv:2406.06606.
37. Kang Y, Zhang Z, & Zhao M, et al. (2022). Tie memories to e-souvenirs: Hybrid tangible AR souvenirs in the museum. *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pp. 1-3.
38. Zhang X, Zhang J, & Rekabdar B, et al. (2024). *Dynamic and adaptive feature generation with LLM*. arXiv preprint arXiv:2406.03505.
39. Zhang X, Zhang J, & Mo F, et al. (2024). *TIFG: Text-informed feature generation with large language models*. arXiv preprint arXiv:2406.11177.
40. Zhang J, Wang X, & Ren W, et al. (2024). *RATT: A thought structure for coherent and correct LLM reasoning*. arXiv preprint arXiv:2406.02746.
41. Wu Z. (2024). Deep learning with improved metaheuristic optimization for traffic flow prediction. *Journal of Computer Science and Technology Studies, 6*(4), 47-53.

ISSN (Online): 2583-2468
Volume-3 Issue-6 || November 2024 || PP. 9-14 DOI: 10.5281/zenodo.14177281

42. Zhang X, Wang Z, & Jiang L, et al. (2024). *TFWT: Tabular feature weighting with transformer*. arXiv preprint arXiv:2405.08403.
43. Wang Z, Chen Y, & Wang F, et al. (2024). *Improved Unet model for brain tumor image segmentation based on ASPP-coordinate attention mechanism*. arXiv preprint arXiv:2409.08588.
44. Tan C, Wang C, & Lin Z, et al. (2024). Editable neural radiance fields convert 2d to 3d furniture texture. *International Journal of Engineering and Management Research*, *14*(3), 62-65.
45. Hu Y, Yang Z, & Cao H, et al. (2020). Multi-modal steganography based on semantic relevancy. *International Workshop on Digital Watermarking*, pp. 3-14. Cham: Springer International Publishing.
46. Wang L, Cheng Y, & Xiang A, et al. (2024). *Application of natural language processing in financial risk detection*. arXiv preprint arXiv:2406.09765.
47. Cheng Y, Yang Q, & Wang L, et al. (2024). *Research on credit risk early warning model of commercial banks based on neural network algorithm*. arXiv preprint arXiv:2405.10762.
48. Xu Q, Wang T, & Cai X. (2024). *Energy market price forecasting and financial technology risk management based on generative AI*.
49. Xiang A, Huang B, & Guo X, et al. (2024). *A neural matrix decomposition recommender system model based on the multimodal large language model*. arXiv preprint arXiv:2407.08942.
50. Hu Y, Cao H, & Yang Z, et al. (2020). Improving text-image matching with adversarial learning and circle loss for multi-modal steganography *International Workshop on Digital Watermarking*, pp. 41-52. Cham: Springer International Publishing.
51. Feng J, Li Y, & Yang Z, et al. (2020). User identity linkage via co-attentive neural network from heterogeneous mobility data. *IEEE Transactions on Knowledge and Data Engineering*, *34*(2), pp. 954-968.