Analysis of the Application of Large Language Models in Three Major Fields: Trading and Portfolio Management, Financial Risk Management, and Financial Text Mining

Hangkai Zhang

School of Economics, Hangzhou Dianzi University Information Engineering College, Hangzhou 311305, China

outlook_57CDCF55613E4C47@outlook.com

Abstract. With the iteration of artificial intelligence, large language models, with their powerful capabilities in natural language understanding, logical reasoning and structured output, are influencing various fields in the financial ecosystem. This article will focus on the application of analytical large language models in three major fields: trading and portfolio management, financial risk management, and financial text mining. In trading and portfolio management, large language models understand market sentiment during the trading and investment process by building different models and can comprehensively and rationally judge whether it is suitable for trading. Through training large language models, future applications are more likely to be a "human-machine combination" model. In terms of financial risk management, different types of large language models have different performance differences. By comparing large language models, more accurate risk analysis can be obtained, and it can help people complete risk management more accurately. In the field of financial text mining, large language models, with their massive structured and unstructured data and their ability to quickly read texts, can obtain more reliable and detailed text reports. This article will systematically review the applications of large language models in three major fields.

Keywords: Trading and Portfolio Management, Financial Risk Management, Financial Text Mining.

1. Introduction

Since its inception, the financial sector has been a field that fully relies on information processing and decision-making. From company financial reports, stock price fluctuations, macroeconomic indicators to social media discussions, a vast amount of structured and unstructured data has jointly influenced financial decision-making and information processing. However, traditional information processing and analysis methods (such as news and reports) often encounter problems of low efficiency, single dimensions, and excessive resource consumption. Moreover, traditional approaches sometimes carry too many personal subjective ideas, which can affect the accuracy of decision-making. In recent years, with the continuous improvement of technological levels, the emergence of artificial intelligence, especially the development of natural language processing technology, new technical methods for processing and analyzing information have emerged. Among them, the Generative AI large language model, with its powerful understanding, generation and reasoning capabilities, influences various fields in the financial ecosystem [2]. A large language model (LLM) is an artificial intelligence model based on deep learning, whose core objective is to understand and infer people's needs. It is characterized by large scale and high complexity, and its performance improves with the growth of parameters and data scale, which is known as the scaling law [9]. This particular model can quickly analyze a large amount of data and conduct multi-dimensional analysis of events, thereby providing more objective ideas.

In the trading and investment industry, large language models can have a strong predictive ability for the stock returns of the day by analyzing news headlines and can quickly collect and analyze complex information to improve market efficiency [8]. In terms of financial risks, the large language model RiskLabs has demonstrated outstanding risk prediction capabilities. Although large language models cannot be used alone for prediction, they can be combined for inference to achieve forward-looking risk analysis [1]. In the field of financial text mining, BloombergGPT can effectively reduce model harm and bias, and improve the accuracy and depth of information processing [11]

This study aims to systematically review the application, challenges, technical approaches and future opportunities of large language models in the three core areas of finance (trading and portfolio management, financial risk management, and financial text mining). This study will systematically review the application of large language models in three major fields through the data and methods of different papers.

2. Trading and Portfolio Management

Large language models have the ability to predict. For instance, the sentiment score generated by GPT-4 based on news headlines can significantly predict the stock returns of the next trading day, thereby generating returns. This predictive ability becomes more pronounced after negative news about small-cap stocks and those with poor liquidity, which is consistent with the theories of arbitrage restrictions and limited investor attention. Meanwhile, this predictive ability is emerging. Only sufficiently complex and advanced LLMs (such as GPT-3.5, GPT-4) demonstrate this ability, while more fundamental and earlier models (such as GPT-1, GPT-2, BERT) cannot predict effectively. And it is vulnerable to being misled by "strategic biases" in corporate press releases (such as misjudgments about stock repurchases and convertible bond issuances). This indicates that there is a "model size threshold" for the predictive ability of large language models (LLMS), and only advanced models with sufficiently large parameter sizes can construct predictions of effective trading portfolios. Moreover, the Sharpe ratios of various large language models on different financial datasets also vary (Figure 1).

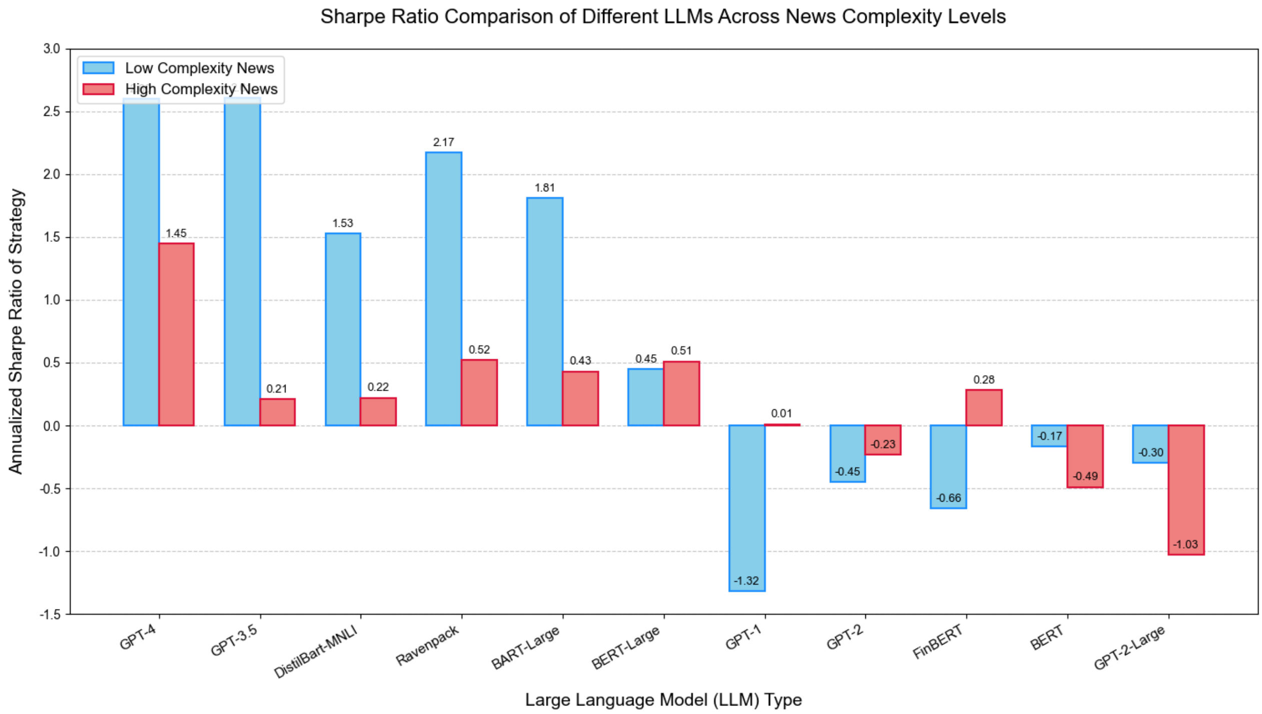

Figure 1. Sharpe ratios of various models.

Data source:[8]

Figure 1 clearly shows the differences in Sharpe ratios of different LLM models in low-complexity and high-complexity news environments. It can be seen that GPT-4 performs well under both types of news complexity. Especially in high-complexity news, its Sharpe ratio far exceeds that of other models, demonstrating a powerful ability to process complex information. However, the basic models such as GPT-1 and GPT-2 have a low Sharpe ratio under high-complexity news, and some are even negative, indicating that they are difficult to effectively handle complex news information.

Meanwhile, LLMs provide dual support of "emotion-driven + data-driven" for trading strategies by analyzing market sentiment, integrating multi-source data, and simulating decision-making logic. It performs particularly well in highly volatile markets such as cryptocurrencies and stocks and has a significant difference in returns compared to traditional strategies (Figure 2).

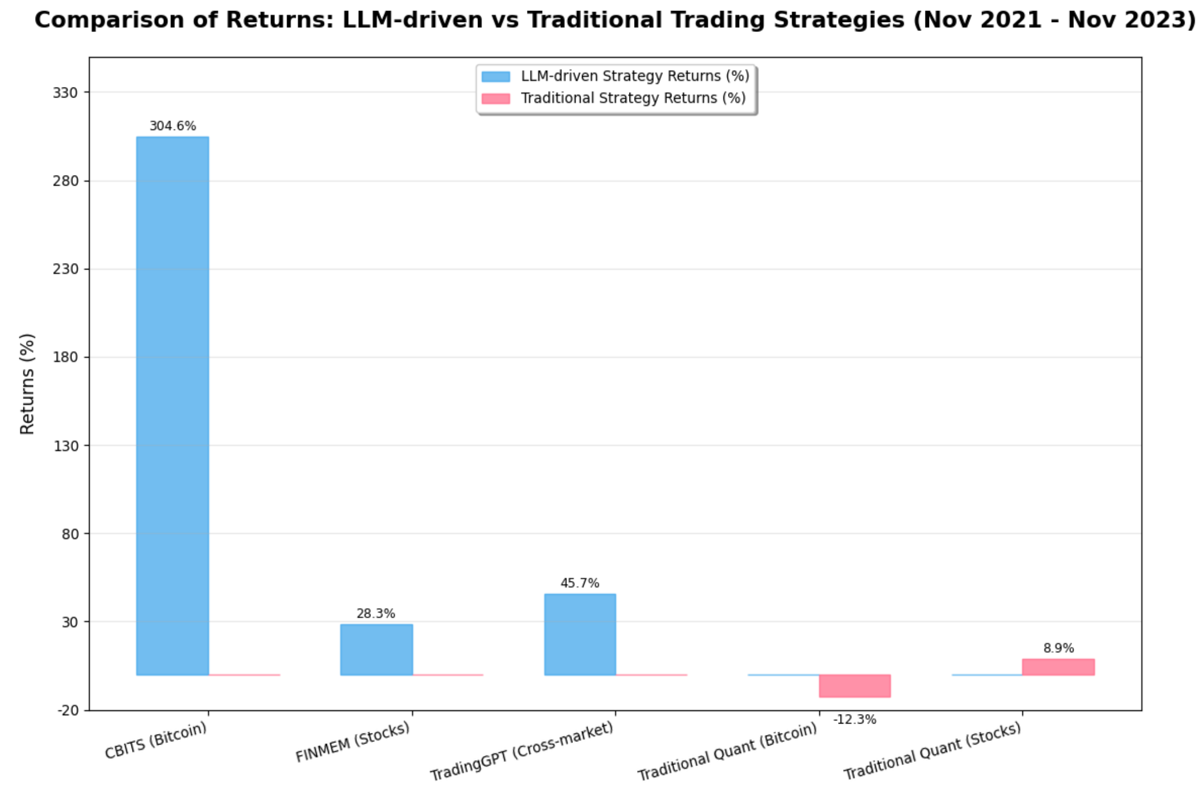

Figure 2. Market return differences.

Data source [7]

Figure 2 visually presents the return advantages of LLM strategies in single-market (Bitcoin, stocks) and cross-market trading, confirming the conclusion in the paper that LLMS enhance trading profitability through sentiment analysis and multi-agent collaboration

However, LLMS vary in different market environments, stock characteristics, and news types. The effectiveness of LLM trading portfolios requires targeted adaptation. For instance, in the stock market: small-cap stocks offer more excess returns. Empirical evidence shows that the prediction coefficient of the ChatGPT 4 strategy for small-cap stocks (NYSE market capitalization below the 10th percentile) (0.102+0.586=0.688) is more than six times that of large-cap stocks (0.102). The reason is that small-cap stocks have stronger "attention constraints", higher arbitrage restrictions, and the market's response to news is more delayed, making the information advantage of LLMS more significant. In terms of the time window dimension, the information absorption cycle is approximately 2 days. The holding period test shows that the excess returns of the ChatGPT 4 strategy are concentrated in the first two trading days after the news release (38bps on the first day and 20bps on the second day), and the returns are no longer significant from the third trading day. It is indicated that the market needs approximately 2 days to fully absorb the news information. The optimal holding period for LLM strategies is 1-2 days. Holding for an excessively long period will erode returns. [4]. Nowadays, LLMS face many significant challenges and deficiencies in this field.

LLMS have a lag in training data. The models used in the paper (such as GPT-4) have their knowledge cut-off dates. They cannot know new events, new companies or macroeconomic changes that occur after the cut-off dates, which will cause their judgments to be based on outdated information.

Slow reasoning speed. The calling speed of the Application Programming Interface of LLM is much slower than that of the traditional quantization signal calculation. In high-frequency trading

(HFT) or scenarios that require millisecond-level responses to news, the current inference speed of LLMS may not meet the requirements.

The decision-making process of LLMS is a black box. It is very difficult to understand which specific keyword or logic in the news it is based on to draw a bullish or bearish conclusion. This has brought huge difficulties to risk management and strategy attribution. If the strategy fails, it is very difficult to determine whether the market style has changed or there is a problem with the model itself.

The decision-making of LLMS requires a huge amount of computing power and incurs high costs. Advanced LLMS such as GPT-4 have great potential in extracting trading signals from text and can serve as an auxiliary and enhanced tool in portfolio management. However, taking it as the core of a fully autonomous trading system is still not mature at present, mainly limited by its lag, black box characteristics, cost, etc. Future applications are more likely to be a "human-machine combination" model, where LLMS are responsible for processing massive amounts of information and providing initial insights, while human investment managers are in charge of final decision-making, risk control, and dealing with extreme market situations that models cannot handle.

3. Regarding financial risk management

Against the backdrop of the rapid development of the digital economy and fintech, financial services are characterized by high frequency, online presence and intelligence. However, the complexity and concealment of financial fraud have also escalated simultaneously. From traditional credit card fraud and identity forgery to new types of telecommunications fraud, false transactions and malicious money laundering, the economic losses caused have been continuously rising, posing a serious threat to the stability of the financial system. Meanwhile, traditional financial risk management methods have limitations [3].

With the deep integration of large language models and financial risk management, LLM, with its advantages of massive sample data and multi-task generalization, has become a key technology for solving financial risk management, providing new paths for dynamic detection and in-depth analysis of fraudulent behavior. LLM safeguards user assets and helps build a safer and more trustworthy digital financial ecosystem. And the core performance of different large language models in financial fraud detection also varies (Figure 3).

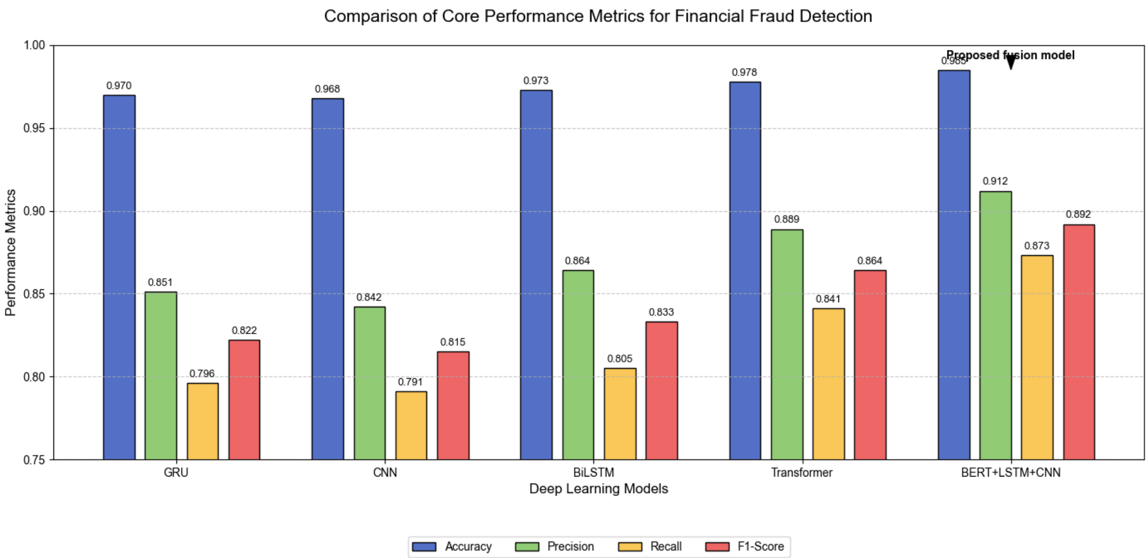

Figure 3. Performance differences

Data source:[3]

As shown in the data in Figure 3, the recall rates of traditional models (GRU, CNN, BiLSTM) are generally low. However, the fusion model, through LLMs semantic embedding and LSTM time series modeling, has increased the recall rate to 0.873, significantly reducing the risk of missed detectives and better meeting the actual needs of financial fraud detection. This technological integration and algorithm optimization contribute to a secure and trustworthy digital financial environment.

In addition, in the lending scenarios of the p2p industry, there is a significant information asymmetry between lenders and borrowers. Lenders often lack sufficient data to assess the credit status of borrowers, and traditional credit scoring models overlook the potential risk information contained in the textual descriptions submitted by borrowers in loan applications (such as the purpose of the loan and personal financial situation explanations). Large language models can be adopted to extract valuable risk information from loans and enhance the accuracy of predictions during the loan approval stage. And the predicted effects also vary in different loan environments (Figure 4).

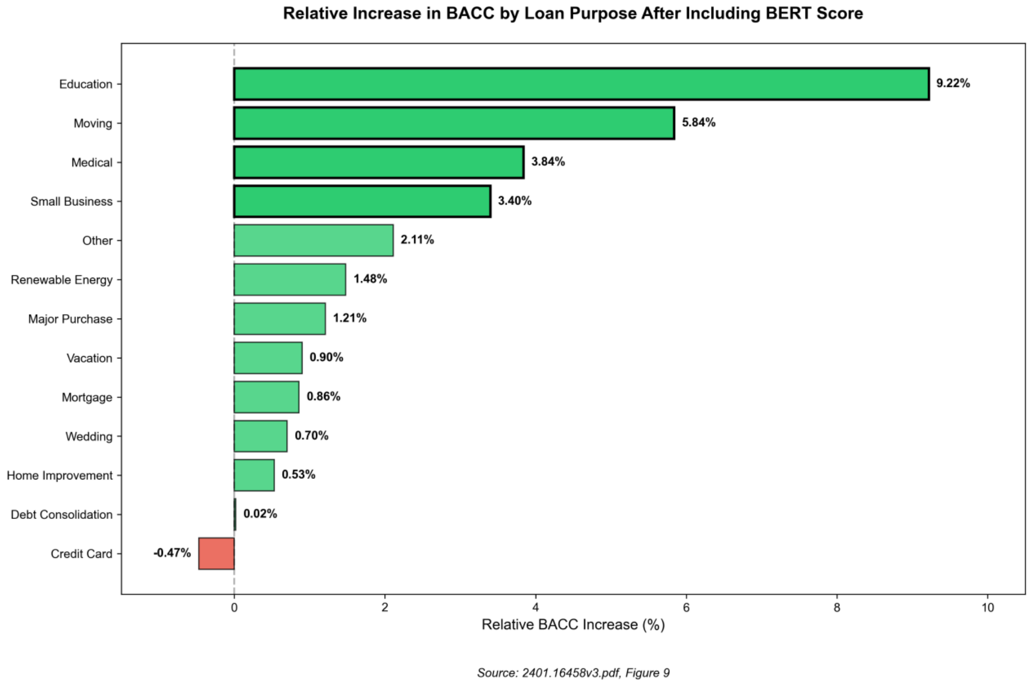

Figure 4. Risk analysis of different loans

Data source :[10]

As shown in Figure 4, this figure visually presents the improvement rate of Balanced Accuracy for various loan purposes. Sorted by the improvement rate from low to high, the BACC improvement of educational loans is the most significant. It reached 9.22%, with significant increases in moving (5.84%), medical (3.84%), and small business loans (3.40%). Credit card loans were the only category that saw a negative increase (-0.47%).

The large language model team innovatively uses a multimodal fusion architecture to uniformly encode and represent structured transaction data, text records, and behavior sequences with BERT, solving the problem of insufficient ability of traditional methods to handle multimodal and unstructured data. To address the imbalance problem, Focal Loss is adopted as the loss function, which effectively alleviates the challenges brought by the imbalance of extreme categories in financial fraud data, significantly improves the recognition ability of minority (fraud) samples, and its performance is superior to the traditional over-sampling method. LLM enhances robustness. This model introduces a contrastive learning mechanism, which enhances the model's ability to distinguish similarities through semantic interference, thereby improving the model's stability and generalization ability in the face of noise interference.

The deep integration of large language models provides an effective technical framework and empirical support for the field of financial risk control, demonstrating its great potential in enhancing the accuracy, robustness and semantic understanding ability of fraud detection, and pointing out the direction for building a more intelligent and reliable financial risk control system. However, at present, all the training data are static and lack practical real applications. There are still challenges in practical applications. In the future, more accurate and forward-looking analyses can be provided for financial risk management from aspects such as enhancing the cognitive level of models and strengthening the understanding of data.

4. Financial Text Mining

Large language models are also widely applied in the field of financial text mining. Large language models (LLMS) based on deep learning and specially pre-trained for the financial field significantly outperform traditional dictionary methods and simpler machine learning models in extracting information from financial texts. Experiments have proved that the large language model FinBERT has outstanding performance in multiple key tasks such as sentiment analysis and ESG topic identification. The large language model FinBERT is a customized large language model for the financial field. This model is based on Google's BERT algorithm and innovatively uses domain-adaptive pre-training. It utilizes a vast financial text database to enable the model to deeply learn relevant professional terms and analysis methods in the financial field (Figure 5).

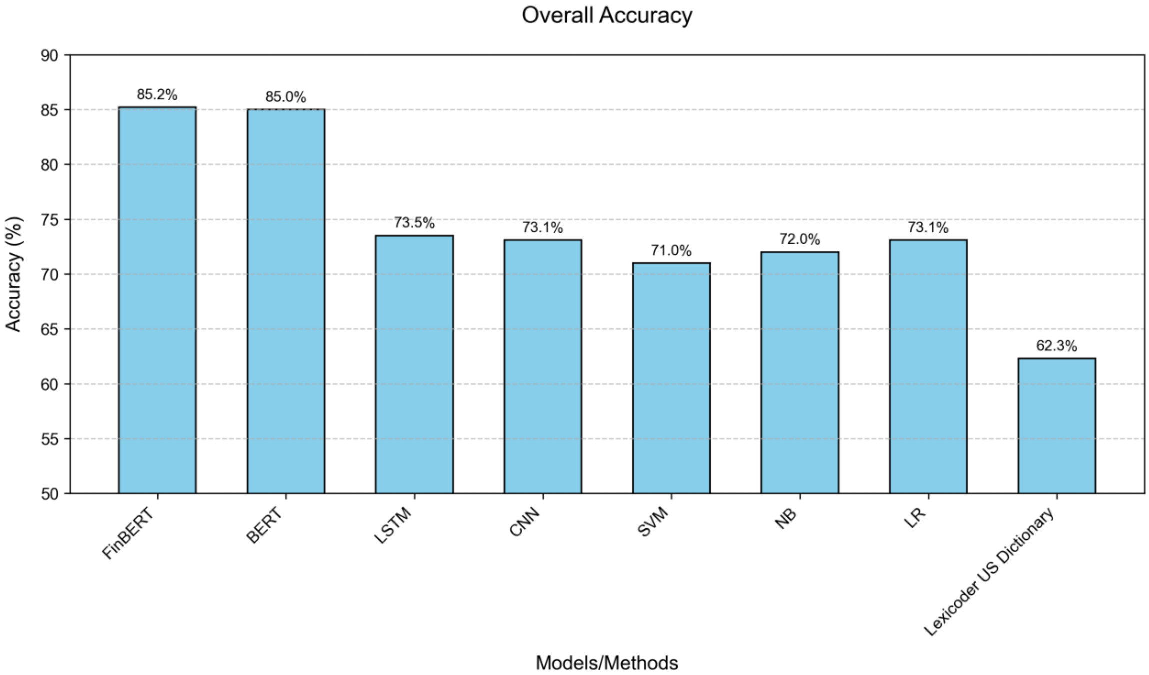

Figure 5. Accuracy

Data source: [5]

Figure 5 clearly shows that FinBERT comprehensively leads all benchmark models in terms of Accuracy, especially in capturing negative emotions, where it has a significant advantage.

The core logic of LLMs in processing financial texts is "first generate high-quality word embeddings and then achieve the analysis goals through fine-tuning downstream tasks." Embedding generation captures the semantic word order of financial texts. Its core is to transform financial texts into dense vectors of different dimensions. Compared with traditional techniques, it can retain word order and context semantics. Downstream task adaptation, from general to financial scenarios, after generating words and embedding them, LLMs ADAPTS to financial tasks through the "pooling layer + fully connected layer". In addition, the core reason why LMs outperforms traditional technologies is that compared with traditional text analysis methods, LLMs have three irreplaceable advantages in financial text processing. The first is the global context understanding ability, which can capture the long-distance dependencies of financial texts. Secondly, there is the transfer learning ability. It can learn comprehensively through a vast amount of text coverage in the pre-training stage and only requires a small amount of text adaptation training in the fine-tuning stage. Furthermore, LLMs can. Imitating human thinking, learning by simulating people's judgments on financial texts to imitate human thinking and generate thought decisions similar to those of humans [6].

The processing capacity for complex financial texts is limited. Deviations in sentiment classification are prone to occur in complex financial texts, and sometimes core emotions cannot be handled. The quality and timeliness of training data are constraints. Large language models are highly dependent on the quality of pre-training data. If there are biases in the training corpus, the model will learn and amplify these defects. At the same time, financial documents are updated too quickly and their timeliness lags behind. The cost of training and deployment is too high, the requirements for hardware are high, and the training time is too long.

Artificial intelligence has provided a new and more powerful tool (FinBERT) for the field of financial text analysis, which can yield more reliable and detailed text analysis results, greatly reducing the need for expensive and time-consuming labeled data and lowering the application threshold. The model can be fine-tuned to be applicable to various tasks, thereby assisting those with relevant needs. The model also fails to truly think like a human when confronted with complex problems, and there are still many aspects that can be improved.

5. Conclusion

Overall, large language models, with their powerful comprehension and analytical capabilities, have become an indispensable part of the three core areas of finance. At the same time, large language models have also driven the transformation of the financial industry from data-driven in the past to intelligence-driven now. However, the current large language models still face many deficiencies and challenges in the three major fields of finance. In the future, large language models in the financial field can develop towards deep integration, security and legality, privacy protection and other aspects.

References

[1] Cao, Y., Chen, Z., Pei, Q., Dimino, F., Ausiello, L., Kumar, P., ... & Ndiaye, P. M. (2024). Risklabs: Predicting financial risk using large language model based on multi-sources data. arXiv preprintarXiv:2404.07452.

[2] Desai, A. P., Ravi, T., Luqman, M., Mallya, G., Kota, N., & Yadav, P. (2024). Opportunities and challenges of generative-AI in finance. In 2024 IEEE International Conference on Big Data (BigData)(pp. 4913-4920). IEEE.

[3] Gong, J., Wang, Y., Xu, W., & Zhang, Y. (2024). A Deep Fusion Framework for Financial Fraud Detection and Early Warning Based on Large Language Models. Journal of Computer Science and Software Applications, 4(8).

[4] Heng, R. Q. W., Vittori, E., Ong, K., Mao, R., Cambria, E., & Mengaldo, G. (2025). Leveraging LLMS for Top-Down Sector Allocation In Automated Trading. arXiv preprintarXiv:2503.09647.

[5] Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. Contemporary Accounting Research, 40(2), 806–841.

[6] Lin, J., & Sun, L. (2025). Text Analysis in Economics and Finance with Large Language Models: Fundamentals, Applications, and Future Prospects. China Journal of Econometrics, 5(1), 1–34.

[7] Liu, C., Arulappan, A., Naha, R., Mahanti, A., Kamruzzaman, J., & Ra, I. H. (2024). Large language models and sentiment analysis in financial markets: A review, datasets and case study. IEEE Access.

[8] Lopez-Lira, A., & Tang, Y. (2023). Can chatgpt forecast stock price movements? return predictability and large language models. arXiv preprintarXiv:2304.07619.

[9] Matarazzo, A., & Torlone, R. (2025). A survey on large language models with some insights on their capabilities and limitations. arXiv preprintarXiv:2501.04040.

[10] Sanz-Guerrero, M., & Arroyo, J. (2024). Credit risk meets large language models: Building a risk indicator from loan descriptions in P2P lending. arXiv preprintarXiv:2401.16458.

[11] Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., ... & Mann, G. (2023). Bloomberggpt: A large language model for finance. arXiv preprintarXiv:2303.17564.