A Comprehensive Review of Vertical Applications in the Financial Sector Based on Large Language Models

Yanlin Mao [2,3,a], Bo Chen[1,b*], Weiqin Chen [2,3,c], Yuandan Deng [2,3,d], Juntao Zeng [2,3,e] and Mengzhen Du[1,f]

[a] 1061143634@qq.com; [b*] bochen@uestc.edu.cn; [c] 2291027652@qq.com; [d] denyuda@163.com; [e] 2448038829@qq.com; [f] 202322090716@std.uestc.edu.cn

[1] School of Information and Software Engineering, University of Electronic Science and Technology of China, Cheng Du, Sichuan, 610000, China
[2] Sichuan Provincial Key Laboratory of Intelligent Terminal jointly built by Hall and City, Yi Bin, Sichuan, 644000, China
[3] University of Electronic Science and Technology of China Yibin Park, Yi Bin, Sichuan, 644000, China

Abstract. This paper investigates the vertical applications of Large Language Models (LLM) in the financial domain, with a focus on the widespread use and profound impact of large language models in the financial industry. It begins by introducing the developmental history of large language models, followed by an overview of mainstream large language models in the financial domain. Subsequently, it explores the applications of large language models in financial text analysis and sentiment analysis, risk management and prediction, as well as customer service and intelligent assistants. Finally, it analyzes the challenges and limitations faced by large language models in the financial sector, proposing potential directions for future technological improvements.

Keywords: Large Language Models, Financial Text Analysis, Sentiment Analysis, Risk Management Prediction, Customer Service, Intelligent Assistant.

1 Introduction

1.1. Background

In November 2022, OpenAI launched ChatGPT, triggering a new wave of technological change and sparking interest and research in Large Language Models (LLM). LLM typically refers to neural network models with parameters numbering in the billions or more[1]. Distinguished by an abundance of parameters and extensive training data, LLMs demonstrate outstanding performance in understanding, analysis, prediction, and question answering, capable of executing a wide range of natural language processing tasks.The rise of this technology, particularly the advent of large language models, has profoundly transformed various aspects of the financial sector. Firstly, leveraging their exceptional natural language processing capabilities, these models enable rapid and accurate processing of vast amounts of financial information, including intelligent analysis of unstructured data and provision of smart customer service. Secondly, in the realm of risk management, large language models can offer more accurate predictions and risk assessments, aiding financial institutions in more effectively managing potential risks. In terms of investment decision-making, these models

BDEDM 2024, January 12-14, Ningbo, People's Republic of China

DOI 10.4108/eai.12-1-2024.2347198

provide powerful tools for analyzing market trends and predicting stock movements, supporting more intelligent investment strategies. On the front of compliance and regulation, the application of large language models can more effectively monitor and address potential compliance risks, enhancing the compliance levels of financial institutions.Overall, large language models are propelling the financial industry towards a digitized, intelligent future, injecting elements of greater intelligence and efficiency into financial services. This article will delve into the vertical applications of large language models in the financial domain, aiming to comprehensively present their applications and far-reaching impacts in the financial sector.

1.2. Rise and development of LLM

Early language models were mainly based on statistical methods, the most typical of which was the n-gram model[2] . These models use frequency statistics to model relationships between words, but are limited by their limited context Windows and ability to capture complex language structures. With the development of deep learning techniques, neural network language models are beginning to attract attention and perform well, such as recurrent neural networks (RNNS) and short term memory networks (LSTMS), which are better able to capture language dependencies over long distances. Word vector models such as Word2Vec and GloVe were also widely used during this period. In 2017, Transformer[3] came out and achieved remarkable success in natural language processing, marking a major leap forward in the field of language models. In 2018, BERT[4] was proposed as a major breakthrough in the direction of pre-trained models for natural language processing. In the same year, the Generative Pre-trained Transformer(GPT-1) [5] was proposed by OpenAI and was the first LLM. By the end of 2022, ChatGPT based on GPT-3 brought the LLM into the public eye, triggering a new round of technological frenzy.

1.3. Structure

This paper is divided into five parts, the structure is as follows:

The first part mainly introduces the background and the development of LLM. The second part gives an overview of the LLM from three aspects: the existing big prediction model, financial data set and evaluation index. In the third part, the application of text analysis and sentiment analysis, risk management and prediction, customer service and intelligent assistant in financial field is introduced in detail. In the fourth chapter, the limitations and challenges of the intelligent question answering task in the financial field are analyzed and discussed. The last part is the summary and prospect of this paper.

2 Overview

2.1. Part of the model

At present, LLM is in its initial stage, and the main research and development comes from the industry, and most of the existing LLM models are general models, and there are relatively few LLM in the financial field. Some LLMS are listed in Table 1.

Table 1. Partial LLM.

Model	Time	Parameter quantity（Billion）
Moss-moon-003-base[6]	2023.02	16
Llama-7B/13B/33B/65B[7]	2023.02	7/13/33/65
BloombergGPT[8]	2023.03	50
XuanYuan[9]	2023.05	10
BaiChuan-7B/13B[10]	2023.06	7/13
Qwen-7B[11]	2023.08	7
DISC-FinLLM[12]	2023.10	13

MOSS is the first open source Chinese LLM to match ChatGPT in training scale and technology. Llama was developed by Meta Corporation and is available in 7B, 13B, 33B and 65B versions. BloombergGPT is the first publicly published LLM in finance. XuanYuan is a model based on BLOOM-176B training. BaiChuan, developed by Baichuan Intelligence, is excellent in both Chinese and English. Qwen-Chat-7B is a large model-based AI assistant developed using an alignment mechanism based on Qwen-7B. DISC-FinLLM comes from FudanDISC and is based on a large financial model fine-tuned by LoRA instructions on Baichuan-13B-Chat. For financial consulting, financial text analysis, financial calculation, financial knowledge retrieval Q&A 4 scenarios.

2.2. Data Set

In Table 2, some of the training datasets commonly used in finance are listed, and in Table 3, some of the test datasets are listed.

Table 2. Train Dataset.

Dataset	Task	data volume	Description
FPB[8]	Sentiment analysis	18690	Financial Phrase Library dataset containing financial news
SmoothNLP NHG[8]	Text generation	4642	Contains a corpus of financial news texts
CCKS2022-event[12]	Text classification	3578	A FEW SHOT event extraction for the financial sector
FDDC2018[12]	Text mining	333	Chinese data sets for tasks such as financial entity identification and relationship extraction.

Table 3. Test Datasets.

Dataset	Description
BBT CFLEB[13]	The professional evaluation data set of financial large models in the Chinese field contains eight standard language tasks, including forum sentiment analysis FinFE, event extraction FinQA, etc
FinEval[14]	Text in the field of finance, covering multiple subtasks such as named entity recognition

FR2KG[15]	It consists of 17,799 entities, 26,798 relational triples, and 1,328 attribute triples, covering 10 entity types, 19 relational types, and 6 attributes.
FLARE_ZH[13]	It covers all aspects of financial natural language processing and financial forecasting
FinanceIQ[13]	Focusing on the Chinese assessment data set in the financial field, it focuses on the assessment of the knowledge and reasoning ability of LLMs in financial scenarios, covering 10 financial categories and 36 financial subcategories, with a total of 7173 single-choice questions.

2.3. valuation index

The evaluation indicators are shown in Table 4.

Table 4. Evaluation indicators.

Index	Description
Accuracy[13]	Represents the proportion of questions answered correctly by the system.
Precision[8]	Represents the percentage of all responses that the system answered correctly
Recall[8]	Represents the percentage of questions answered correctly by the system out of all actual correct questions.
F1[13]	Is a harmonic mean of accuracy and recall, used to combine the two metrics.
BLEU[13]	Evaluates the quality of generative responses.
ROUGE[13]	Evaluate the overlap between the abstracts generated by the system and the reference abstracts.
MAP[14]	All candidate answers are scored and sorted according to the size of the score. The higher the correct answer is, the larger the MAP value is. The maximum value is 1.
MRR[13]	Measures the quality of the first correct result returned by the system.
NDCG[13]	evaluate the quality of sorting results.

3 Application of Large Language Models in the Financial Sector

3.1. Text Analysis and Sentiment Analysis

Large language models perform text and sentiment analysis in the financial domain by leveraging deep learning to analyze extensive financial texts, supporting investment decisions, and understanding market sentiment. Araci [16] et al pioneered the application of BERT in the financial sector. They utilized two financial sentiment analysis datasets to pre-train FinBERT and further fine-tune it for sentiment analysis tasks on financial news articles, laying the foundation for financial BERT. Yi Yang [17] et al pre-trained a BERT model specific to the financial domain using a massive financial communication corpus. Experimental results on three financial sentiment classification tasks demonstrated the superiority of the model over general-purpose BERT, providing a pre-trained model for natural language processing tasks in the financial field. Xiao-Yang Liu [18] et al introduced the FinGPT framework, applying the Low-rank Adaptation (LoRA) method to reduce trainable model parameters. They also incorporated Reinforcement Learning with Stock Prices (RLSP), using relative changes in stock prices as output labels to guide model training, enabling cost-effective development of large financial language models for sentiment analysis in quantitative trading. Jiangtong Li [19] et al constructed an annotated sentiment dataset using social media posts, research reports, and investment ratings. They introduced CFGPT, enhancing the model's understanding and reasoning abilities through continuous training and supervised fine-tuning, achieving financial sentiment analysis. Yi Yang [20] et al proposed InvestLM, utilizing a curated dataset of financial investment-related instructions. Through fine-tuning a large language model with small yet high-quality instructional data, the model demonstrated outstanding performance in text comprehension and investment advice.

3.2. Risk Management and Prediction

Large language models contribute to enhanced risk management and prediction in the financial sector by leveraging deep learning and natural language processing techniques to analyze extensive textual data, improving the accuracy of risk identification and market trend forecasting. Pawan Kumar Rajpoot [21] et al have employed GPT for financial relationship extraction on the EFinD dataset, with GPT4 based on EPR demonstrating excellent performance, providing technological support for financial risk management and prediction. Bhaskarjit Sarmah [22] et al propose a comprehensive approach, integrating retrieval-enhanced LLMs and metadata to enhance the reliability and accuracy of information extraction from financial profit reports. Ziao Wang [23] et al innovatively introduce the FinVis-GPT, a multimodal large language model designed for understanding financial charts, accomplishing various tasks related to financial charts through pre-training and instruction fine-tuning. Yang Li [24] et al present the TraingGPT trading agent framework, enhancing automated trading results significantly through collaborative interactions among multi-agent systems and addressing limitations in GPT memory processing.

3.3. Customer Service and Intelligent Assistants

In the financial sector, large language models support customer service and intelligent assistant functions through advanced natural language processing technology, fostering a more intelligent and efficient financial service experience. Shi Yu [25] et al have developed a chatbot based on the deep bidirectional transformer (BERT) model for customer service in financial investment, equipped with the ability to recognize 381 intents. Yanbo J. Wang [26] et al leverage the pre-trained language model De-BERTa, employing optimization methods such as multi-model fusion and training set combinations to implement a reasoning program for financial report generation. Karmvir Singh Phgat [27] et al addressing the sensitivity of large language models to few-shot examples, employ zero-shot prompts for complex domain-specific numerical reasoning. Xuanyu Zhang [28] et al introduce XuanYuan2.0, integrating both general and domain-specific knowledge, providing accurate and contextually relevant responses for the Chinese financial domain through pre-training and fine-tuning. The Summary of Papers on Text and Sentiment Analysis Using Financial Large Language Models in Table 5.

Table 5. Summary of Papers on Text and Sentiment Analysis Using Financial Large Language Models.

Paper	Advantages	Paper	Advantages
Paper[16]	Pioneered the application of BERT in the financial domain	Paper[23]	Shows advantages in financial chart-related tasks, surpassing existing state-of-the-art multimodal LLMs.
Paper[17]	Provided a pre-trained financial-specific language model	Paper[24]	Marks the first appearance of an LLM agent trading system with integrated role design.
Paper[18]	Introduced an open-source data framework, addressing the issue of limited financial dataset	Paper[25]	Effectively enhances the handling of escalation issues, reducing intent misclassification.
Paper[19]	Proposed fine-tuning strategy, enhancing model's understanding and reasoning abilities	Paper[26]	Achieves an execution accuracy of 68.99, program accuracy of 64.53, and secures the first position in the FinQA challenge.
Paper [20]	Introduced InvestLM financial large language model with exceptional understanding capabilities	Paper [27]	Designs zero-shot prompts effectively extracting complex domain-specific numerical inferences.
Paper[21]	Demonstrates good performance in extracting valuable information from financial documents.	Paper[28]	Strong knowledge base and conversational capabilities.
Paper[22]	Addresses issues of irrelevant entity detection in QA systems with multiple documents, enhancing system reliability and accuracy.		

4 lication of Large Language Models in the Financial Sector

Due to the excellent performance of large language models in automation and transferability, they demonstrate strong universality and high accuracy in the financial sector. However, these models also face inevitable drawbacks and challenges, including ethical considerations and potential biases that may arise in the financial decision-making process. In financial decision-making, the model's predictions can significantly impact the fairness and equity of decisions. If the model training data contains biases, it may introduce unfair elements into the decision-making process. Therefore, ensuring the fairness and diversity of data is crucial. Additionally, decision-makers need to be mindful of the potential impact of the model on different demographic groups to ensure social responsibility in decision-making. When implementing data privacy measures, such as anonymization and encryption, a balance between security and privacy must be struck to mitigate reidentification risks and additional computational costs. In terms of transparency and interpretability, model interpretation techniques like LIME and SHAP can be employed to help decision-makers understand the logic behind the model's decisions, but the limitations of these techniques should be acknowledged. Lastly, regarding adversarial sample training, cautious measures need to be taken to prevent threats to the model's security and stability, while simultaneously balancing the negative impact on accuracy and generalization.

5 Conclusions

Large language models play a crucial role in the financial sector, providing real-time decision support, risk management, and personalized services, bringing substantial value to the industry. These models have the capability to analyze vast amounts of data in real-time and find widespread applications in intelligent risk management, personalized financial services, and market forecasting. Key areas of technological innovation include defense against adversarial sample attacks, enhancing model interpretability, privacy protection techniques, and multi-model integration. Defense against adversarial sample attacks involves exploring methods based on adversarial training to strengthen model robustness. Improving interpretability can be achieved by developing explanation models based on attention mechanisms. In the realm of privacy protection techniques, attention should be given to the development of federated learning and secure multi-party computation technologies. Additionally, multi-model integration contributes to enhancing the overall robustness of decision-making systems. These directions hold the potential to bring new possibilities to financial technology**.**

Acknowledgments. Part of it is supported by Yibin city science and technology plan project: DZKJDX2021020004. This paper is supported by Sichuan Province Sci-Tech Project Grant: 2022YFG0167. Part of it is funded by Sichuan Province Sci-Tech Project Grant: 2023YFG0032.

References

[1] Carlini N, Tramer F, Wallace E, et al. Extracting training data from large language models[C //30th USENIX Security Symposium (USENIX Security 21). 2021: 2633-2650.

[2] Stahlberg F. Neural machine translation: A review[J]. Journal of Artificial Intelligence Research, 2020, 69: 343-418.

[3] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[4] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

[5] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J]. 2018.

[6] T Sun, X Zhang, Z He, P Li, Q Cheng, H Yan, X Liu, Y Shao, Q Tang, X Zhao, et al. Moss: Training conversational language models from synthetic data. 2023.

[7] Touvron H, Lavril T, Izacard G, et al. Llama: Open and efficient foundation language models[OL . arXiv preprint arXiv:2302.13971, 2023.

[8] Wu S, Irsoy O, Lu S, et al. Bloomberggpt: A large language model for finance[J]. arXiv preprint arXiv:2303.17564, 2023.

[9] Zhang X, Yang Q. Xuanyuan 2.0: A large chinese financial chat model with hundreds of billions parameters[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023: 4435-4439.

[10] Baichuan-inc. Baichuan-13b. https://github.com/baichuan-inc/Baichuan-13B, 2023a.
[11] Qwen. Qwen-7b. https://github.com/QwenLM/Qwen-7B, 2023.
[12] Chen W, Wang Q, Long Z, et al. DISC-FinLLM: A Chinese Financial Large Language Model based on Multiple Experts Fine-tuning[J]. arXiv preprint arXiv:2310.15205, 2023.
[13] Chang Y, Wang X, Wang J, et al. A survey on evaluation of large language models[J]. arXiv preprint arXiv:2307.03109, 2023.
[14] Zhang L, Cai W, Liu Z, et al. Fineval: A chinese financial domain knowledge evaluation benchmark for large language models[J]. arXiv preprint arXiv:2308.09975, 2023.
[15] Wang, W.G., et al.: Data set and evaluation of automated construction of financial knowledge graph. Data Intelligence 3(3), 418-443 (2021). doi: 10.1162/dint_a_00108
[16] Araci, Dogu. "Finbert: Financial sentiment analysis with pre-trained language models." arXiv preprint arXiv:1908.10063 (2019).
[17] Yang, Yi, Mark Christopher Siy Uy, and Allen Huang. "Finbert: A pretrained language model for financial communications." arXiv preprint arXiv:2006.08097 (2020).
[18] Liu,Xiao-Yang, Guoxuan Wang, and Daochen Zha. "Fingpt: Democratizing internet-scale data for financial large language models." arXiv preprint arXiv:2307.10485 (2023).
[19] Li, Jiangtong, et al. "CFGPT: Chinese Financial Assistant with Large Language Model." arXiv preprint arXiv:2309.10654 (2023).
[20] Yang, Yi, Yixuan Tang, and Kar Yan Tam. "Investlm: A large language model for investment using financial domain instruction tuning." arXiv preprint arXiv:2309.13064 (2023).
[21] Rajpoot P K, Parikh A. Gpt-finre: In-context learning for financial relation extraction using large language models[J]. arXiv preprint arXiv:2306.17519, 2023.
[22] Sarmah B, Zhu T, Mehta D, et al. Towards reducing hallucination in extracting information from financial reports using Large Language Models[J]. arXiv preprint arXiv:2310.10760, 2023.
[23] Wang Z, Li Y, Wu J, et al. FinVis-GPT: A Multimodal Large Language Model for Financial Chart Analysis[J]. arXiv preprint arXiv:2308.01430, 2023.
[24] Li Y, Yu Y, Li H, et al. TradingGPT: Multi-Agent System with Layered Memory and Distinct Characters for Enhanced Financial Trading Performance[J]. arXiv preprint arXiv:2309.03736, 2023.
[25] Yu S, Chen Y, Zaidi H. AVA: A financial service chatbot based on deep bidirectional transformers[J]. Frontiers in Applied Mathematics and Statistics, 2021, 7: 604842.
[26] Wang Y J, Li Y, Qin H, et al. A Novel DeBERTa-based Model for Financial Question Answering Task[J]. arXiv preprint arXiv:2207.05875, 2022.
[27] Phogat K S, Harsha C, Dasaratha S, et al. Zero-Shot Question Answering over Financial.
[28] Zhang X, Yang Q. Xuanyuan 2.0: A large chinese financial chat model with hundreds of billions parameters[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023: 4435-4439.