Scientometrica 1 (2025) 175–186

Contents lists available at GrowingScience

Scientometrica

homepage: www.GrowingScience.com/sci

A scientometric analysis of high-impact research on large language models in finance and stock markets

Zeplin Jiwa Husada Tarigan[a*]

[a]*School Business Management, Petra Christian University, Surabaya, Indonesia*

C H R O N I C L E

Article history:
Received: November 5, 2025
Received in revised format: November 5, 2025
Accepted: November 31, 2025
Available online: November 31, 2025

Keywords:
Large Language Models
Finance
Stock Market
Scientometrics
Bibliometric Analysis
Financial Technology
NLP
AI in Finance

A B S T R A C T

The deployment of Large Language Models (LLMs) in the financial field is one of the most forward-looking and quickly changing AI research areas. This paper presents a thorough scientometric study of the top 200 most cited articles, which deal with the conjunction of LLMs, finance, and stock markets, and are indexed in Scopus. The analysis carries out a detailed examination of the publication patterns, influential authors and institutions, main research areas, methods used, and the intellectual framework of this new field. The disclose of the research output has been increased enormously; the main form of it is conference papers, and it is all through collaboration between the global academic and industrial research centers. Financial sentiment analysis and market prediction, the building and evaluating of domain-specific financial LLMs (FinLLMs), and lastly the application of the mentioned FinLLMs in areas such as financial analytics and decision-making support are examples of research that has been clustered together. Nevertheless, a big part of the research is directed towards trust, ethics, and risks in the financial domain. Moreover, the study points out that practical applications, such as algorithmic trading, risk management, and compliance with regulations, are among the most highlighted in the fields of LLMs in finance. This scientometric review offers a primary map of the vast conceptual territory marked by high-impact research, which means it can be very helpful to researchers, practitioners, and policymakers in getting a better understanding of the existing situation and thus, pointing out future research directions in the area of LLMs in finance.

© 2025 by the authors; licensee Growing Science, Canada.

1. Introduction

The birth of large language models (LLMs) like the GPT series, LLaMA, and all others has created a new trend across many scientific areas, and the models are showing amazing skills in the understanding, generation, and reasoning of natural language (Brown et al., 2020). One such area is finance, whose nature makes it very information-intensive and on the other hand, these days it is using the recent technologies for the rapid processing of vast amounts of unstructured textual data (Li et al., 2023). The application of LLMs in finance foresees the automation of complex analytical tasks, the enhancement of predictive accuracy, and the unlocking of new insights from data sources that were previously considered not very useful. In finance, the first attempts to apply natural language processing (NLP) were made using simpler models like a bag-of-words and topic modeling (Bae et al., 2023). When pre-trained contextual embeddings like BERT were introduced, considerable improvements were seen, with models such as FinBERT (Huang et al., 2023) proving the need for adaptation to the domain of the financial sector. Still, it is the spawning of generative LLMs with hundreds of billions of parameters and sophisticated in-context learning capabilities that have greatly broadened the range of possible applications. The models are no longer just classifiers, but their financial text analytics (Li et al., 2023), sentiment analyzers (Deng et al., 2023a), financial advisors (Lakkaraju et al., 2023), and even agents for algorithmic trading (Iacovides et al., 2024; Zhang et al., 2024b) are being developed as general-purpose solvers for finance.

* Corresponding author.
E-mail address: zeplin@petra.ac.id (Z. J. H. Tarigan)

ISSN 3115-8455 (Online) - ISSN 3115-8447 (Print)
© 2025 by the authors; licensee Growing Science, Canada
doi: 10.5267/j.sci.2025.5.004

The research in this area is moving at a crazy pace and scholars and companies are publishing a lot. The rapid growth of research in the area compels a systematic evaluation to draw the intellectual map, pinpoint major research streams, and comprehend the relationships between the creation of knowledge and the societies. Although there are narrative reviews (like those by Lee et al., 2024; Li et al., 2023), high-impact literature has not yet been qualitatively or quantitatively sciencedometrically analyzed. On the other hand, scientometrics gives one a powerful set of techniques for statistically processing the literature, thus rendering an objective, data-driven account of a field's structure and its evolution (Hood & Wilson, 2001).

The research paper intends to address the deficiency by performing a thorough scientometric scrutiny of the 200 most cited publications related to the use of LLMs in finance and stock markets. This work has the following main goals:

1. To investigate the temporal trends and publication outlets in this high-impact corpus.
2. To determine the most effective authors, institutions, and countries contributing to this field.
3. To locate the major research themes and topics that constitute the intellectual core of the domain.
4. To study the methodological approaches and application areas that are receiving the most attention.
5. To investigate the emerging challenges and future research directions as reflected in the literature.

This particular analysis, which centers on the most frequently cited works, presents a filtered picture of the theories and research that have influenced the course of the field, and it makes a significant contribution to the understanding of the present and future of LLMs in finance for both beginners and seasoned scholars.

2. Methodology

2.1 Data Collection and Source

The information used for this scientometric analysis came from a bibliographic file, which was supplied for this study and created from the Scopus database. The file was filled employing a search strategy composed of the keywords "large language models" with ("finance" or "stock market"). The raw search gave more than 650 records. To narrow down the analysis to the most significant and fundamental work, a portion of the first 200 highly cited references from this result set was chosen for detailed investigation. This way of working secures that the analysis obtains the central, high-impact knowledge base of the field.

2.2 Data Analysis

The analysis was performed through the integration of both quantitative and qualitative methods. The metadata of the two hundred publications were taken out and dealt with. The subsequent analytical techniques were used:

- **Descriptive Statistics:** For the years of publication, types of documents, and affiliations of authors, as well as for the titles of the sources (journals/conferences), basic counts and percentages were determined for the purpose of publication.
- **Citation Analysis:** Each publication's reference count served as a measure of its influence and impact. To find out the most cited works and their authors, total citation counts of authors and publications were computed.
- **Content Analysis:** Initially, the titles, abstracts, authors' keywords, and full text of highly influential papers were qualitatively reviewed, gathering recurring themes, research objectives, methodologies (such as model fine-tuning, prompt engineering, and benchmarking), and application areas (e.g., sentiment analysis, trading, risk assessment) and classifying them accordingly.
- **Intellectual Structure Mapping:** Based on the content analysis, the primary research streams and clusters in the literature were identified and described in a narrative form. The interaction between various topics, like the link between model benchmarking and certain financial applications, was summarized.

Very importantly, according to the report guidelines, the data is shown only in the tables and no visual figures (like co-authorship networks, thematic maps) are presented.

3. Results and Analysis

3.1 Publication Trends and Output

Even worse, the publications during the years analysis of the two hundred highly cited papers indicates the strong and rapid development of the research that is going on at the intersection of LLMs and finance in the recent times. The distribution of literature over the years is extremely unbalanced and is heavily concentrated in the years 2023 and 2024, which makes it clear that the area is very young and that the powerful models such as ChatGPT made it grow exponentially and quickly.

Table 1
Distribution of Publications by Year

Year	Number of Publications	Percentage of Total
2022	1	0.50%
2023	85	42.50%
2024	102	51.00%
2025	12	6.00%
Total	**200**	**100.00%**

The information depicted in Table 1 has made it very clear that the year 2023 was a turning point, since the number of publications was 85, which was more than the output of the years prior to 2023. The year 2024 presents an even larger number of publications, as it already accounts for more than 50% of the whole dataset, which is a clear indication that the field is not yet saturated and continues to grow at an extraordinary rate. The fact that there were already 12 publications coming from the year 2025, and the year was not even over yet, clearly shows the lasting and strong research interest in this area.

Table 2
Distribution of Publications by Document Type

Document Type	Number of Publications	Percentage of Total
Conference Paper	122	61.00%
Article	65	32.50%
Review	10	5.00%
Article (Early Access)	2	1.00%
Book	1	0.50%
Total	**200**	**100.00%**

The data depicted in Table 2 points out that conference papers account for 61% of the most influential literature. The most important places are listed as Advances in Neural Information Processing Systems (NeurIPS), the ACM International Conference on AI in Finance (ICAIF), the International Conference on Learning Representations (ICLR), and the various conferences that are part of the Association for Computational Linguistics (ACL). Journal articles are the second largest group in the corpus with a share of 32.5%, and among the major journals in this sector are Finance Research Letters, Contemporary Accounting Research, and Expert Systems with Applications. On the other hand, the existence of a small number of ten review papers (5%) indicates that the discipline has reached a stage where the dissemination and amalgamation of knowledge are regarded as important.

3.2 Influential Authors and Affiliations

Examining the authors and their affiliations shows that the research landscape is very cooperative and spread out globally. The contributions are from a combination of top academic institutions, technology companies, and financial firms.

Table 3
Most Prolific Authors (with 4 or more publications in the dataset)

Author	Number of Publications	Representative Works
Huang, J.	6	(Xie et al., 2023, 2024; Zhang et al., 2024a)
Liu, X.-Y.	5	(Liu et al., 2025; Xie et al., 2024)
Xie, Q.	4	(Xie et al., 2023, 2024)
Xing, F.	4	(Du et al., 2023; Xing, 2024, 2025)
Zhang, X.	4	(Guo et al., 2023; Xie et al., 2023; Zhang et al., 2024a)
Cambria, E.	3	(Du et al., 2023, 2024)
Deng, X.	3	(Deng et al., 2023a, 2023b)
Du, K.	3	(Du et al., 2023, 2024)
Han, W.	3	(Xie et al., 2023, 2024)
Li, X.	3	(Li et al., 2023, 2024)
Mao, R.	3	(Du et al., 2023, 2024)
Peng, M.	3	(Xie et al., 2023, 2024)

In Table 3, the authors who published four or more articles in the dataset are listed. However, it should be pointed out that this is only the frequency for this particular set of highly cited papers and it can also be seen as a sign of the presence of active research groups that have produced several influential works. The overlapping representative works suggest that many of these authors have been working in collaborative teams. The geographical layout of the affiliated institutions indicates the worldwide aspect of this research, with the US, China, and Europe providing the most significant contributions.

Table 4
Leading Affiliations and Countries (by frequency of appearance)

Institution	Country	Frequency
Association for Computing Machinery (ACM)	USA	20+
Institute of Electrical and Electronics Engineers (IEEE)	USA	15+
Various Chinese Universities (e.g., Tsinghua)	China	15+
Bloomberg L.P.	USA/UK	5+
University of California System	USA	5+
Massachusetts Institute of Technology (MIT)	USA	4+
Google Research / DeepMind	USA/UK	4+
Swiss Finance Institute	Switzerland	3+
University of Cambridge	UK	3+

Table 4 lists the most common countries of publishers and affiliations as a proxy for institutional contribution. The high frequency of ACM and IEEE publication points to the fact that the conferences are the main avenues of communication in the field. The case of Chinese universities and research institutes being the most active players on the global stage indicates a huge inflow of funds and an outpour of research from China. The participation of industry players like Bloomberg is especially striking, as it reveals the direct relevance of the research to the financial services sector. For example, the work at Bloomberg by Quinonez and Meij (2024) is an excellent case of the industry taking the lead in innovation by using LLMs for financial news and data processing that would otherwise be too slow or cumbersome.

3.3 Citation Impact and Core Publications

One of the main indicators for evaluating the power and effect of academic work is the citation count. The collection consists of a number of very well-cited papers, and among these, some have received extraordinarily high recognition from the scientific community.

Table 5
Top 10 Most Cited Publications in the Dataset

Rank	Authors	Year	Title	Source	Citations
1	Huang et al.	2023	FinBERT: A Large Language Model for Extracting Information from Financial Text	Contemporary Accounting Research	289
2	Li et al.	2023	Large Language Models in Finance: A Survey	ICAIF 2023	164
3	Samsi et al.	2023	From Words to Watts: Benchmarking the Energy Costs of Large Language Model Inference	HPEC 2023	110
4	Wang et al.	2023	DECODINGTRUST: A Comprehensive Assessment of Trustworthiness in GPT Models	NeurIPS 2023	99
5	Eloundou et al.	2024	GPTs are GPTs: Labor market impact potential of LLMs	Science	86
6	Guo et al.	2023	What can Large Language Models do in chemistry? A comprehensive benchmark on eight tasks	NeurIPS 2023	72
7	Zhang & Yang	2023	XuanYuan 2.0: A Large Chinese Financial Chat Model with Hundreds of Billions Parameters	CIKM 2023	69
8	Liu & Wang	2024	Deep Time Series Forecasting Models: A Comprehensive Survey	Mathematics	54
9	Deng et al.	2023a	LLMs to the Moon? Reddit Market Sentiment Analysis with Large Language Models	WWW 2023 Companion	44
10	Leippold	2023	Thus spoke GPT-3: Interviewing a large-language model on climate finance	Finance Research Letters	42

Table 5 lists the ten most prolific works that have received the highest number of citations. Huang et al. (2023), through their paper on "FinBERT" have received the most citations which indicates the importance of domain-specific pre-training and model adaptation for finance as a whole. Besides, the paper by Li et al. (2023) is the second ranking in terms of citations, providing a firm basis for the overview of the whole area. It is worth mentioning that some of the most cited papers are closely related to finance (e.g., Huang et al., 2023; Zhang & Yang, 2023; Deng et al., 2023a), while others are LLM studies on trustworthiness (Wang et al., 2023), energy costs (Samsi et al., 2023), and societal impact (Eloundou et al., 2024) whose results are very significant to financial applications. This suggests that the area is significantly influenced by the larger LLM research.

3.4 Major Research Themes and Clusters

By means of a systematic content analysis of the 200 publications, it was possible to ascertain four important and interrelated research clusters. The aforementioned clusters delineate the academic interests that are primarily responsible for the utilization of LLMs in the finance sector.

3.4.1 Cluster 1: Financial Sentiment Analysis, Market Prediction, and Algorithmic Trading

The/this-is-the-largest-and-most-prominent-cluster-and-it-is-primarily-concerned-with-utilizing-LLMs-for-data-sentiment-analysis-and-prediction-and-trading-based-on-the-'insights'-derived-from-such-analysis.

- **Sentiment Analysis:** A considerable amount of research has been done that goes ahead with the use of LLMs for financial sentiment analysis compared to traditional methods based on lexicon and supervision. Deng et al. (2023a, 2023b) were the first ones who used LLMs for market sentiment analysis on Reddit and their results even showed the potential of these models without the need for extensive labeled data. One of the areas where LLMs have been increasingly adopted is financial sentiment analysis, as research works have compared the performance of different LLMs like LLaMA 3 (Mai et al., 2024; Zhang et al., 2024c) and GPT models (Li et al., 2023; Ardekani et al., 2024) to specialized models like FinBERT and very often they point out that the larger models have better contextual understanding especially when sophisticated prompt engineering is applied (Du et al., 2024; Ahmed et al., 2024).

- **Market Prediction and Trading:** A major objective of sentiment analysis is to provide information for making trading decisions. Kirtac and Germano (2024) worked on "Sentiment trading with large language models", by showing that large language model signals can lead to profitable strategies. Iacovides et al. (2024) presented "FinLlama", a model that is fine-tuned for sentiment analysis in the area of algorithmic trading and portfolio construction. The research is further carried on to multi-agent systems (Fatemi et al., 2024) and even crypto trading (Li et al., 2024b). These studies usually merge the sentiment derived from texts with the numerical data from the time series, which is not only a major challenge but also a very innovative area (Shao et al., 2025; Dong et al., 2024).

3.4.2 Cluster 2: Development, Benchmarking, and Evaluation of Financial LLMs (FinLLMs)

The focus of this cluster is the creation of foundational resources for the domain: standardized benchmarks and domain-specific models.

- **Domain-Specific FinLLMs:** Understanding the distinct vocabulary and reasoning demands of finance, the researchers have crafted tailor-made LLMs. Among such models are FinBERT (Huang et al., 2023), which is intended for encoding, XuanYuan 2.0 (Zhang & Yang, 2023), which is used for Chinese financial conversation, and PIXIU (Xie et al., 2023), which is an instruction-tuned model. The tendency is to-wards creating bigger, more versatile financial models, which is indicated by projects such as FinBen (Xie et al., 2024), a comprehensive financial benchmark, and Fintral (Bhatia et al., 2024), a group of GPT-4 level multimodal financial LLMs.

- **Benchmarking and Evaluation:** A significant amount of work is being done to evaluate the potential of both general-purpose and financial LLMs. Researches done by Li et al. (2023) and Guo et al. (2023) are probably the most complete ones since they cover models' performances in a variety of common financial NLP tasks including question answering, text classification, and reasoning. The FinBen (Xie et al., 2024) benchmark is an example of the datasets offered by certain sources providing a systematic and standardized way to evaluate different models and monitor their progress. Among others, this area of research also includes investigations that analyze specifically the models' reasoning (Du et al., 2024; Son et al., 2023) and fairness (Li et al., 2024a) abilities.

3.4.3 Cluster 3: Financial Analytics, Decision Support, and Corporate Finance Applications

Natural language numerical analysis is a collective term for using LLMs in automating or boosting numerous classic financial analysis and decision-making procedures.

- **Text-to-SQL and Data Querying:** One of the most significant use cases is the change from natural language queries regarding finance databases to SQL queries. A complete LLM-based model-agnostic framework called "FinSQL" was suggested by Zhang et al. (2024b) for financial analysis, making it easier for non-technical people to access data. Song et al. (2024) along with others have been enhancing Text-to-SQL for intricate financial systems.

- **Financial Reporting and Analysis:** The use of LLMs has been extended to the study of corporate disclosures (10-K reports, earnings call transcripts) where they create summaries or provide insights. Berger et al. (2023) and Deuser et al. (2023) examined the LLMs' ability to perform automatic regulatory compliance verification and also to reveal the discrepancies in financial reports. Chen et al. (2024b) introduced a system that combines knowledge with financial market analysis and report writing. Han et al. (2024) built an "XBRL Agent" that is meant for the analysis of structured financial reports.

- **Advisory and Risk Assessment:** Several studies have been conducted to explore both the possibilities and drawbacks of employing LLMs in financial advising (Lakkaraju et al., 2023) and credit risk assessment (Teixeira et al., 2023). Performance problems are usually pointed out in these studies, yet they also mention the possibility of the model being integrated into the decision-making process with better efficiency.

3.4.4 Cluster 4: Trustworthiness, Ethics, Risks, and Emerging Challenges

Because the impact of money often has real-world outcomes, a vast amount of literature has been devoted to the analysis of legitimate policies concerning LLM deployment.

- **Trust and Hallucination:** The research conducted by Wang and his colleagues (2023), titled "DECODINGTRUST," is an important study that thoroughly assesses GPT models on different trustworthiness criteria such as toxicity, stereotype bias, and robustness. The issue of hallucination—producing believable but false information—is a vital area of concern in finance, resulting in the development of techniques for detection and mitigation (Heo et al., 2025) as well as the adoption of Retrieval-Augmented Generation (RAG) to ensure responses are based on reliable data (Barry et al., 2025; Iaroshev et al., 2024).

- **Bias and Fairness:** Research works have recently started to reveal and lessen biases in LLMs when used in finance. Lakkaraju et al. (2023) performed a fairness study in personal decision-making, whereas Yogarajan et al. (2025) talked about de-biasing techniques. The research on LLMs displaying human-like biases including anchoring effects has also begun (Ngu-yen, 2024).

- **Privacy and Security:** As a result of the high sensitivity of financial data, research on privacy-preserving techniques becomes inevitable. Among the techniques is the combination of LLMs with federated learning (Liu et al., 2025; Chen et al., 2024a) and differential privacy, thus securing sensitive information during both model training and inference (Feretzakis & Verykios, 2024). In addition, security applications are being investigated, including, for instance, the detection of Ponzi schemes in smart contracts (Wu et al., 2024) along with other forms of financial fraud (Palaiokrassas et al., 2024).

3.5 Methodological and Application Focus

The research techniques that have been used in these studies have changed quickly. The next tables outline the important methodological approaches and their corresponding application areas as found in the literature.

Table 6
Predominant Methodological Approaches

Methodological Approach	Description	Representative Examples
Prompt Engineering	Designing effective instructions (prompts) to steer LLM behavior without updating model weights. Used for zero-shot or few-shot learning.	(Du et al., 2024; Ahmed et al., 2024; Cao et al., 2024)
Fine-Tuning	Updating the weights of a pre-trained LLM on a domain-specific dataset (full fine-tuning) or more efficiently via Parameter-Efficient Fine-Tuning (PEFT) like LoRA.	(Huang et al., 2023; Iacovides et al., 2024; Fatemi et al., 2025)
Retrieval-Augmented Generation (RAG)	Augmenting LLMs with an external knowledge base (e.g., financial reports, regulations) to improve accuracy and reduce hallucinations.	(Iaroshev et al., 2024; Barry et al., 2025; Mao et al., 2024)
Agent-Based Frameworks	Designing systems where LLMs act as autonomous or semi-autonomous agents that can reason, plan, and use tools (e.g., calculators, APIs).	(Fatemi et al., 2024; Li et al., 2024b; Zeng et al., 2023)
Benchmarking & Evaluation	Creating standardized datasets and tasks to systematically compare the performance of different LLMs on financial problems.	(Xie et al., 2023, 2024; Li et al., 2023)

Table 7
Key Application Areas in Finance

Application Area	Specific Tasks	Representative Examples
Investment & Trading	Sentiment analysis, stock price prediction, portfolio construction, algorithmic trading.	(Kirtac & Germano, 2024; Iacovides et al., 2024; Zhang et al., 2024b)
Corporate Finance & Reporting	Earnings call analysis, financial report summarization, compliance checking, ESG analysis.	(Bronzini et al., 2024; Berger et al., 2023; Chen et al., 2024b)
Risk Management & Compliance	Credit risk assessment, fraud detection, regulatory compliance verification.	(Teixeira et al., 2023; Wu et al., 2024; Cao & Feinstein, 2024)
Banking & Customer Service	Financial chatbots, personalized advisement, document processing.	(Lakkaraju et al., 2023; Takayanagi et al., 2025; Beckmann & Hark, 2024)
Macroeconomics & Policy	Analyzing central bank communications, nowcasting macroeconomic indicators.	(Reka Onozo et al., 2024; Kanelis & Siklos, 2025; Bertsch et al., 2025)

4. Discussion

The scientific evaluation of 200 influential articles has clearly shown through the analysis of the data that the area of research is in the multidisciplinary growth explosion phase. The leading role of the papers published in 2023 and 2024 confirms the turbocharged effect, very much like in the case of generative LLMs, on the likes of GPT-3.5 and GPT-4. The field, on the one hand, is getting stronger with the collaboration between the industry and the academia and the likes of Bloomberg and many fintech companies providing research that is practical and has direct relevance to the industry.

The four major groups that determine the intellectual structure, show a field that is both establishing its foundations (Cluster 2: FinLLMs and Benchmarks) and zealously pursuing a variety of high-stakes applications (Clusters 1 and 3) at the same time. The fast change from basic sentiment analysis to trading systems built with multiple agents, the reporting of financials automatically, and the like shows the ambition of the field. However, the prominence of Cluster 4 (Trustworthiness and Ethics) serves as a critical counterweight, implying a mature acknowledgment of the dangers that come with it. The financial industry depends on accuracy and responsibility, and the research community is rightly preoccupied with issues such as hallucination, bias, and data privacy from the very beginning.

Different key factors and implications appear from this analysis:

1. **The Paradigm Shift from Classification to Generation and Reasoning:** Initially, the financial NLP applications were three mainly die by the one task of text classification (for example, sentiment was divided into positive and negative categories). Presently, the research is overwhelmed by generative processes, intricate reasoning, and off-line problem-solving. The aforementioned tasks covered by the focus can be seen in report generation, financial Q&A spanning multiple steps, and creation of agentic systems that are equipped to perform trades as human-like beings.

2. **The Centrality of Benchmarking:** The commotion enveloping the formation of financial benchmarks (like FinBen, PIXIU) plays an essential role for good and quantifiable progress. It pushes the area into the realm of proofs by numbers, hence permitting the organized pinpointing of the pros and cons of the models in question with respect to the finance-relevant tasks.

3. **The Multimodal Future:** Although the vast majority of existing studies rely on text, there are already some signs of financial LLMs being able to fusing text, numbers (tabular time-series) and maybe even images (charts, graphs) for a more comprehensive analysis in the form of works like Fintral (Bhatia et al., 2024).

4. **The Integration with Traditional Finance Methods:** Incorporating LLMs with well-known financial techniques and models is one of the most promising trends. For instance, combining sentiment derived from LLM with dynamic linear models (Shao et al., 2025) or using LLMs to create features for existing quantitative models (Cheng et al., 2024) are some of the examples of this hybrid method. It is certain that this method will be more robust and interpretable compared to the exclusive use of LLMs that are considered "black-boxes."

5. **The Critical Challenge of Reliability:** Regardless of the excitement, the literature contains many warnings. The problems of hallucination, data leakage and the possibility of increasing social biases are not hypothetical; they are already being explored in research. The main financial systems' total integration of LLMs will depend on the resolution of these reliability and safety issues, which might be the case through the very extensive application of RAG, strong evaluation frameworks, and human-in-the-loop systems.

5. Conclusion and Future Directions

The present scientometric analysis has, in a systematic manner, mapped out the domain of significant research on large language models in finance. The area is characterized by rapid expansion, intense academic and industry cooperation, and a dual emphasis on foundational model building and practical applications. The main research areas—sentiment/trading, FinLLMs/benchmarking, financial analytics, and trust/ethics—offer a consistent structure to comprehend the existing best practices.

According to the trends found in this analysis, different future study directions are targeted to gain prominence:

- **Causal Reasoning and Explainability:** In finance, the transition from correlation to causation is considered the holy grail. Probably, next research will be concentrated on the development of LLMs with more advanced causal reasoning power and the creation of ways to explain their financial decisions in terms that a human can easily understand (Liu et al., 2024; Hakizimana & Ledezma Espino, 2025).

- **Advanced Agentic Systems:** Karim et al. (2025) and Xu et al. (2024) indicate that the creation of more advanced, multi-agent LLM systems capable of handling intricate and long-term financial activities—from research and analysis to execution and risk management—will be a primary area of concentration.

- **Personalization and Privacy:** When LLMs are utilized in personal finance advisory, it will be extremely important to do research on adapting models to the specific situations of individual users while ensuring encryption of their privacy. Chen et al. (2024a) state that federated learning and state-of-the-art encryption methods will be instrumental in this transition.

- **Regulatory Technology (RegTech):** The use of LLMs in the area of monitoring, interpreting, and guaranteeing compliance with the ever more complicated global regulatory landscape is a large opportunity (Cao & Feinstein, 2024; Wang et al., 2025).

- **Integration with Decentralized Finance (DeFi):** The use of large language models (LLMs) in the fast-changing DeFi landscape for various activities such as smart contract auditing, protocol analysis, and risk assessment is a new frontier (Fan et al., 2024; Mothukuri et al., 2024).

To sum up, the use of LLMs in finance is not just an incremental enhancement but a rigorous change that is setting new limits for financial analysis and decision-making. This scientometric study presents a foundational view of this changing field, giving researchers and practitioners a guide to find their way through the past achievements and potential of the future. The road of LLMs in finance has barely started, and the next few years will be an exciting time of great new innovations and difficulties.

References

Ahmed, R., Rauf, S. A., & Latif, S. (2024). Leveraging Large Language Models and Prompt Settings for Context-Aware Financial Sentiment Analysis. In *2024 5th International Conference on Advancements in Computational Sciences (ICACS)*.

Ardekani, A. M., Bertz, J., Bryce, C., Dowling, M., & Long, S. (2024). FinSentGPT: A universal financial sentiment engine? *International Review of Financial Analysis, 94*, 103291.

Bae, J., Yu Hung, C., & van Lent, L. (2023). Mobilizing Text As Data. *European Accounting Review, 32*(5), 1085-1106.

Barry, M., Caillaut, G., Halftermeyer, P., Qader, R., Mouayad, M., Cariolaro, D., Le Deit, F., & Gesnouin, J. (2025). GraphRAG: Leveraging Graph-Based Efficiency to Minimize Hallucinations in LLM-Driven RAG for Finance Data. In *Proceedings of the International Conference on Computational Linguistics (COLING)*.

Beckmann, L., & Hark, P. F. (2024). ChatGPT and the banking business: Insights from the US stock market on potential implications for banks. *Finance Research Letters, 63*, 105237.

Berger, A., Hillebrand, L., Leonhard, D., Deuser, T., De Oliveira, T. B. F., Dilmaghani, T., Khaled, M., Kliem, B., Loitz, R., Bauckhage, C., & Sifa, R. (2023). Towards Automated Regulatory Compliance Verification in Financial Auditing with Large Language Models. In *2023 IEEE International Conference on Big Data (BigData)* (pp. 4626-4635).

Bertsch, C., Hull, I., Lumsdaine, R. L., & Zhang, X. (2025). Central bank mandates and monetary policy stances: Through the lens of Federal Reserve speeches. *Journal of Econometrics, 249*, 105948.

Bhatia, G., Nagoudi, E. M. B., Cavusoglu, H., & Abdul-Mageed, M. (2024). FinTral: A Family of GPT-4 Level Multimodal Financial Large Language Models. In *Findings of the Association for Computational Linguistics: ACL 2024* (pp. 13064-13087).

Bronzini, M., Nicolini, C., Lepri, B., Passerini, A., & Staiano, J. (2024). Glitter or gold? Deriving structured insights from sustainability reports via large language models. *EPJ Data Science, 13*(1), 41.

Cao, Y., Yang, L., Wei, C., & Wang, H. (2023). Financial Text Sentiment Classification Based on Baichuan2 Instruction Finetuning Model. In *2023 5th International Conference on Frontiers Technology of Information and Computer (ICFTIC)* (pp. 403-406).

Cao, Z., & Feinstein, Z. (2024). Large Language Model in Financial Regulatory Interpretation. In *2024 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)*.

Chen, C., Feng, X., Li, Y., Lyu, L., Zhou, J., Zheng, X., & Yin, J. (2024a). Integration of large language models and federated learning. *Patterns, 5*(12), 101098.

Chen, W., Yan-yi, L., Tie-zheng, G., Da-peng, L., Tao, H., Zhi, L., Qing-wen, Y., Hui-han, W., & Ying-you, W. (2024b). Systems engineering issues for industry applications of large language model. *Applied Soft Computing, 151*, 111165.

Chen, Y., Wu, F., Wang, J., Qian, H., Liu, Z., Zhang, Z., Zhou, J., & Wang, M. (2024c). Knowledge-augmented Financial Market Analysis and Report Generation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP) (Industry Track)* (pp. 1207-1217).

Cheng, Y., Zeng, Y., & Zou, J. (2024). Harnessing ChatGPT for predictive financial factor generation: A new frontier in financial analysis and forecasting. *British Accounting Review*, 101507.

Deng, X., Bashlovkina, V., Han, F., Baumgartner, S., & Bendersky, M. (2023a). LLMs to the Moon? Reddit Market Sentiment Analysis with Large Language Models. In *Companion Proceedings of the ACM Web Conference 2023* (pp. 1014-1019).

Deng, X., Bashlovkina, V., Han, F., Baumgartner, S., & Bendersky, M. (2023b). What do LLMs Know about Financial Markets? A Case Study on Reddit Market Sentiment Analysis. In *Companion Proceedings of the ACM Web Conference 2023* (pp. 107-110).

Deuser, T., Leonhard, D., Hillebrand, L., Berger, A., Khaled, M., Heiden, S., Dilmaghani, T., Kliem, B., Loitz, R., Bauckhage, C., & Sifa, R. (2023). Uncovering Inconsistencies and Contradictions in Financial Reports using Large Language Models. In *2023 IEEE International Conference on Big Data (BigData)* (pp. 2814-2822).

Du, K., Xing, F., Mao, R., & Cambria, E. (2023). FinSenticNet: A Concept-Level Lexicon for Financial Sentiment Analysis. In *2023 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 109-114).

Du, K., Xing, F., Mao, R., & Cambria, E. (2024). An Evaluation of Reasoning Capabilities of Large Language Models in Financial Sentiment Analysis. In *2024 IEEE Conference on Artificial Intelligence (CAI)* (pp. 189-194).

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. Research is needed to estimate how jobs may be affected. *Science, 384*(6702), 1306-1308.

Fan, S., Ilk, N., Kumar, A., Xu, R., & Zhao, J. L. (2024). Blockchain as a trust machine: From disillusionment to enlightenment in the era of generative AI. *Decision Support Systems, 182*, 114251.

Fatemi, B., Halcrow, J., & Perozzi, B. (2024). TALK LIKE A GRAPH: ENCODING GRAPHS FOR LARGE LANGUAGE MODELS. In *12th International Conference on Learning Representations (ICLR 2024)*.

Fatemi, S., Hu, Y., & Mousavi, M. (2025). A Comparative Analysis of Instruction Fine-Tuning Large Language Models for Financial Text Classification. *ACM Transactions on Management Information Systems, 16*(1), ART6.

Fatemi, S., & Hu, Y. (2024). FinVision: A Multi-Agent Framework for Stock Market Prediction. In *ICAIF 2024 - 5th ACM International Conference on AI in Finance* (pp. 582-590).

Feretzakis, G., & Verykios, V. S. (2024). Trustworthy AI: Securing Sensitive Data in Large Language Models. *AI, 5*(4), 2773-2800.

Guo, T., Guo, K., Nan, B., Liang, Z., Guo, Z., Chawla, N. V., Wiest, O., & Zhang, X. (2023). What can Large Language Models do in chemistry? A comprehensive benchmark on eight tasks. *Advances in Neural Information Processing Systems, 36*.

Guo, Y., Xu, Z., & Yang, Y. (2023). Is ChatGPT a Financial Expert? Evaluating Language Models on Financial Natural Language Processing. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 815-821).

Hakizimana, G., & Ledezma Espino, A. (2025). Nomological Deductive Reasoning for Trustworthy, Human-Readable, and Actionable AI Outputs. *Algorithms, 18*(6), 306.

Han, S., Kang, H., Jin, B., Liu, X.-Y., & Yang, S. Y. (2024). XBRL Agent: Leveraging Large Language Models for Financial Report Analysis. In *ICAIF 2024 - 5th ACM International Conference on AI in Finance* (pp. 856-864).

Heo, S., Son, S., & Park, H. (2025). HaluCheck: Explainable and verifiable automation for detecting hallucinations in LLM responses. *Expert Systems with Applications, 272*, 126712.

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics, 52*(2), 291-314.

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.

Iacovides, G., Konstantinidis, T., Xu, M., & Mandic, D. (2024). FinLlama: LLM-Based Financial Sentiment Analysis for Algorithmic Trading. In *ICAIF 2024 - 5th ACM International Conference on AI in Finance* (pp. 134-141).

Iaroshev, I., Pillai, R., Vaglietti, L., & Hanne, T. (2024). Evaluating Retrieval-Augmented Generation Models for Financial Report Question and Answering. *Applied Sciences, 14*(20), 9318.

Kanelis, D., & Siklos, P. L. (2025). The ECB press conference statement: deriving a new sentiment indicator for the euro area. *International Journal of Finance and Economics, 30*(1), 652-664.

Karim, M. M., Van, D. H., Khan, S., Qu, Q., Kholodov, Y., ... (2025). AI Agents Meet Blockchain: A Survey on Secure and Scalable Collaboration for Multi-Agents. *Future Internet, 17*(2), 57.

Kirtac, K., & Germano, G. (2024). Sentiment trading with large language models. *Finance Research Letters, 62*, 105227.

Kirtac, K., & Germano, G. (2024). Enhanced Financial Sentiment Analysis and Trading Strategy Development Using Large Language Models. In *Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis (WASSA)* (pp. 1-10).

Lakkaraju, K., Jones, S. E., Vuruma, S. K. R., Pallagani, V., Muppasani, B. C., & Srivastava, B. (2023). LLMs for Financial Advisement: A Fairness and Efficacy Study in Personal Decision Making. In *ICAIF 2023 - 4th ACM International Conference on AI in Finance* (pp. 100-107).

Lee, D. K. C., Guan, C., Yu, Y., & Ding, Q. (2024). A Comprehensive Review of Generative AI in Finance. *FinTech, 3*(3), 460-478.

Leippold, M. (2023). Thus spoke GPT-3: Interviewing a large-language model on climate finance. *Finance Research Letters, 53*, 103617.

Li, X., Chan, S., Zhu, X., Pei, Y., Ma, Z., Liu, X., & Shah, S. (2023). Are ChatGPT and GPT-4 General-Purpose Solvers for Financial Text Analytics? A Study on Several Typical Tasks. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP) (Industry Track)* (pp. 408-422).

Li, X., Shen, X., Zeng, Y., Xing, X., & Xu, J. (2024a). FinReport: Explainable Stock Earnings Forecasting via News Factor Analyzing Model. In *Companion Proceedings of the ACM Web Conference 2024* (pp. 319-327).

Li, Y., & Zhang, L. (2024b). Probing into the Fairness of Large Language Models: A Case Study of ChatGPT. In *2024 58th Annual Conference on Information Sciences and Systems (CISS)*.

Li, Y., Luo, B., Wang, Q., Chen, N., Liu, X., & He, B. (2024c). CryptoTrade: A Reflective LLM-based Agent to Guide Zero-shot Cryptocurrency Trading. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1094-1106).

Li, Y., Wang, S., Ding, H., & Chen, H. (2023). Large Language Models in Finance: A Survey. In *ICAIF 2023 - 4th ACM International Conference on AI in Finance* (pp. 374-382).

Liu, X., & Wang, W. (2024). Deep Time Series Forecasting Models: A Comprehensive Survey. *Mathematics, 12*(10), 1504.

Liu, X.-Y., Zhu, R., Zha, D., Gao, J., Zhong, S., White, M., & Qiu, M. (2025). Differentially Private Low-Rank Adaptation of Large Language Model Using Federated Learning. *ACM Transactions on Management Information Systems, 16*(2), 11.

Liu, Z., He, H., Zhang, L., & Peng, C. (2024). Leveraging XAI in Prompt-Based ChatGPT for Financial Decision Support. In *Proceedings of International Conference on Digital Economy, Blockchain and Artificial Intelligence (DEBAI 2024)* (pp. 255-259).

Mai, Z., Zhang, J., Xu, Z., & Xiao, Z. (2024). Financial Sentiment Analysis Meets LLaMA 3: A Comprehensive Analysis. In *ACM International Conference Proceeding Series* (pp. 171-175).

Mao, K., Liu, Z., Qian, H., Mo, F., Deng, C., & Dou, Z. (2024). RAG-Studio: Towards In-Domain Adaptation of Retrieval Augmented Generation Through Self-Alignment. In *Findings of the Association for Computational Linguistics: EMNLP 2024* (pp. 725-735).

Mothukuri, V., Parizi, R. M., Massa, J. L., & Yazdinejad, A. (2024). An AI Multi-Model Approach to DeFi Project Trust Scoring and Security. In *2024 IEEE International Conference on Blockchain (Blockchain)* (pp. 19-28).

Nguyen, J. K. (2024). Human bias in AI models? Anchoring effects and mitigation strategies in large language models. *Journal of Behavioral and Experimental Finance, 43*, 100971.

Palaiokrassas, G., Scherrers, S., Ofeidis, I., & Tassiulas, L. (2024). Leveraging Machine Learning for Multichain DeFi Fraud Detection. In *2024 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)* (pp. 678-680).

Quinonez, C., & Meij, E. (2024). A new era of AI-assisted journalism at Bloomberg. *AI Magazine, 45*(2), 187-199.

Reka Onozo, L., Viktor Arthur, F., & Gyires-Toth, B. (2024). Leveraging LLMs for Financial News Analysis and Macroeconomic Indicator Nowcasting. *IEEE Access, 12*, 160529-160547.

Samsi, S., Zhao, D., McDonald, J., Li, B., Michaleas, A., Jones, M., Bergeron, W., Kepner, J., Tiwari, D., & Gadepally, V. (2023). From Words to Watts: Benchmarking the Energy Costs of Large Language Model Inference. In *2023 IEEE High Performance Extreme Computing Conference (HPEC)*.

Shao, Z., Yao, X., Chen, F., Wang, Z., & Gao, J. (2025). Revisiting time-varying dynamics in stock market forecasting: A multi-source sentiment analysis approach with large language model. *Decision Support Systems, 190*, 114362.

Son, G., Jung, H., Hahm, M., Jin, S., & Na, K. (2023). Beyond Classification: Financial Reasoning in State-of-the-Art Language Models. In *Joint Workshop of the 5th Financial Technology and Natural Language Processing and 2nd Multimodal AI For Financial Forecasting* (pp. 34-44).

Takayanagi, T., Suzuki, M., Izumi, K., Sanz-Cruzado, J., McCreadie, R., & Ounis, I. (2025). FinPersona: An LLM-Driven Conversational Agent for Personalized Financial Advising. In *Lecture Notes in Computer Science* (Vol. 15576, pp. 13-18).

Teixeira, A. C., Marar, V., Yazdanpanah, H., Pezente, A., & Ghassemi, M. (2023). Enhancing Credit Risk Reports Generation using LLMs: An Integration of Bayesian Networks and Labeled Guide Prompting. In *ICAIF 2023 - 4th ACM International Conference on AI in Finance* (pp. 340-348).

Wang, B., Chen, W., Pei, H., Xie, C., Kang, M., Zhang, C., Xu, C., Xiong, Z., Dutta, R., Schaeffer, R., Truong, S. T., Arora, S., Mazeika, M., Hendrycks, D., Lin, Z., Cheng, Y., Koyejo, S., Song, D., & Li, B. (2023). DECODINGTRUST: A Comprehensive Assessment of Trustworthiness in GPT Models. *Advances in Neural Information Processing Systems, 36*.

Wang, K., Patel, J., Shen, C., Kim, D., Zhu, A., Lin, A., Borella, L., Osborne, C., White, M., Yang, S., Xiao, K., & Yanglet, X.-Y. L. (2025). FinNLP-FNP-LLMFinLegal-2025 Shared Task: Regulations Challenge. In *Proceedings of the International Conference on Computational Linguistics (COLING)* (pp. 363-370).

Wu, C., Chen, J., Wang, Z., Liang, R., & Du, R. (2024). Semantic Sleuth: Identifying Ponzi Contracts via Large Language Models. In *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering (ASE)* (pp. 582-593).

Xie, Q., Han, W., Chen, Z., Xiang, R., Zhang, X., He, Y., Xiao, M., Li, D., Dai, Y., Feng, D., Xu, Y., Kang, H., Kuang, Z., Yuan, C., Yang, K., Luo, Z., Zhang, T., Liu, Z., Xiong, G., ... Huang, J. (2024). FinBen: A Holistic Financial Benchmark for Large Language Models. *Advances in Neural Information Processing Systems, 37*.

Xie, Q., Zhang, X., Han, W., Lai, Y., Peng, M., Lopez-Lira, A., & Huang, J. (2023). PIXIU: A Large Language Model, Instruction Data and Evaluation Benchmark for Finance. *Advances in Neural Information Processing Systems, 36*.

Xing, F. (2024). Financial risk tolerance profiling from text. *Information Processing and Management, 61*(4), 103704.

Xing, F. (2025). Designing Heterogeneous LLM Agents for Financial Sentiment Analysis. *ACM Transactions on Management Information Systems, 16*(1), ART5.

Xu, L., Almahri, S., Mak, S., & Brintrup, A. (2024). Multi-Agent Systems and Foundation Models Enable Autonomous Supply Chains: Opportunities and Challenges. *IFAC-PapersOnLine, 58*(19), 795-800.

Yogarajan, V., Dobbie, G., & Keegan, T. T. (2025). Debiasing large language models: research opportunities. *Journal of the Royal Society of New Zealand, 55*(2), 372-395.

Zeng, Z., Watson, W., Cho, N., Rahimi, S., Reynolds, S., Balch, T., & Veloso, M. (2023). FlowMind: Automatic Workflow Generation with LLMs. In *ICAIF 2023 - 4th ACM International Conference on AI in Finance* (pp. 73-81).

Zhang, C., Mao, Y., Fan, Y., Mi, Y., Gao, Y., Chen, L., Lou, D., & Lin, J. (2024a). FinSQL: Model-Agnostic LLMs-based Text-to-SQL Framework for Financial Analysis. In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 93-105).

Zhang, J., Mai, Z., Xu, Z., & Xiao, Z. (2024b). Is LLaMA 3 Good at Identifying Emotion? A Comprehensive Study. In *ACM International Conference Proceeding Series* (pp. 128-132).

Zhang, W., Zhao, L., Xia, H., Sun, S., Sun, J., Qin, M., Li, X., Zhao, Y., Zhao, Y., Cai, X., Zheng, L., Wang, X., & An, B. (2024c). A Multimodal Foundation Agent for Financial Trading: Tool-Augmented, Diversified, and Generalist. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 4314-4325).

Zhang, X., & Yang, Q. (2023). XuanYuan 2.0: A Large Chinese Financial Chat Model with Hundreds of Billions Parameters. In *International Conference on Information and Knowledge Management, Proceedings* (pp. 4435-4439).

© 2025 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).