Information Retrieval in Finance: Industry and Academic Perspectives on Innovation

Chung-Chi Chen
c.c.chen@acm.org
National Institute of Advanced Industrial Science and Technology
Japan

Yongjae Lee
yongjaelee@unist.ac.kr
Ulsan National Institute of Science and Technology
South Korea

Alejandro Lopez-Lira
alejandro.lopez-lira@warrington.ufl.edu
University of Florida
United States

Chanyeol Choi
jacobchoi@linqalpha.com
LinqAlpha
United States

Richard Mccreadie
Richard.Mccreadie@glasgow.ac.uk
University of Glasgow
United Kingdom

Javier Sanz-Cruzado
javier.sanz-cruzadopuig@glasgow.ac.uk
University of Glasgow
United Kingdom

Abstract

Information retrieval (IR) plays a critical role in financial decision-making across investment research, trading, risk management, and reporting. With the rise of large language models (LLMs), IR systems have evolved to support more natural, context-aware workflows. In this tutorial, we survey recent advances in applying IR and LLM technologies in finance, covering agent-based simulations, investor recommender systems, retrieval-augmented research management, and LLM-driven portfolio construction. We highlight practical challenges and propose future research directions at the intersection of IR, LLMs, and financial innovation. More materials can be found at http://irfin.nlpfin.com/.

CCS Concepts

• **Information systems → Information systems applications**.

Keywords

Agentic AI, recommender system, retrieval-augmented generation, portfolio selection

ACM Reference Format:
Chung-Chi Chen, Yongjae Lee, Alejandro Lopez-Lira, Chanyeol Choi, Richard Mccreadie, and Javier Sanz-Cruzado. 2025. Information Retrieval in Finance: Industry and Academic Perspectives on Innovation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), July 13–18, 2025, Padua, Italy.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3726302.3731685

1 Introduction

Financial professionals leverage information retrieval (IR) technology across nearly every facet of the industry. IR systems enable efficient information access and decision support across a variety of use cases. In asset management and investment research, portfolio managers and analysts use IR to retrieve insights from news, filings, and market data. Modern systems allow natural language queries to access relevant information instantly, improving analysis speed and depth. Quantitative trading relies on IR to extract signals from unstructured data like news, social media, and earnings call transcripts. Analysts scan text for patterns such as merger rumors or policy changes and search historical data for similar past events. Risk management and compliance depend on IR to process regulatory texts, risk reports, and audits. For example, IR can retrieve relevant clauses from lengthy banking regulations or detect red flags in communications related to fraud or insider trading. Results must be traceable for audit and regulatory scrutiny. Credit analysis and underwriting involve collecting data from financial reports, credit scores, and industry news. NLP enables accurate fact extraction, while knowledge graphs help uncover relationships—like supplier networks or legal exposure—that affect creditworthiness. IR is also vital in sourcing ESG data scattered across reports, news, and social media. Analysts extract metrics such as emissions or diversity and track controversies. Given evolving ESG language, retrieval combines keyword search with AI-driven semantic understanding. In financial reporting, IR helps parse and extract insights from large volumes of reports and filings. Analysts use natural language search to locate details like lease obligations. Some firms deploy chatbots that summarize earnings or find legal risks, backed by citations. IR tools enhance decision-making by providing fast, accurate access to financial data. By spanning applications such as investment research, trading, risk and compliance, credit and ESG analysis, and financial reporting, it is evident that IR technology is ubiquitous in the financial sector. Each domain imposes distinct demands—such as real-time processing in trading or high recall and traceability in compliance—yet the common objective remains: to transform unstructured information into actionable intelligence. However, due to the broad scope of the topic, this paper primarily focuses on investment-related applications.

The advent of large language models (LLMs) has marked a transformative shift in the way unstructured information is processed and understood. Beyond achieving substantial performance gains across a wide spectrum of natural language processing tasks, LLMs have fundamentally reshaped how users engage with IR systems—ushering in more natural, conversational, and context-aware interactions. Against this backdrop, where IR and NLP are becoming

SIGIR '25, Padua, Italy

ACM ISBN 979-8-4007-1592-1/2025/07
https://doi.org/10.1145/3726302.3731685

increasingly integrated, this paper investigates the evolving role of IR in the financial domain through the lens of investment decision-making. We examine four pivotal areas: agent-based simulations for modeling investment behavior, personalized recommender systems tailored for investors, LLM-driven approaches to portfolio construction, and retrieval-augmented generation (RAG) systems designed for managing investment research workflows. Through this exploration, we aim not only to map the current landscape but also to offer insights into the emerging directions and opportunities at the intersection of LLMs, IR, and financial decision-making.

2 Agent-based Investment Decision Simulation

2.1 Recent Developments

The rise of LLMs has accelerated the use of natural language (NL) as the primary interface for Agent-based Modeling (ABM), shifting from rule-based agents to more cognitively plausible ones capable of rich reasoning and interaction. LLM-based agents have been applied across various domains, from simulating individual and social behavior to coordinating complex workflows. These simulations often use NL prompts and multi-turn interactions to emulate human-like decision-making, with evaluation metrics including task success rates and communication coherence.

In finance, agent-based simulations serve descriptive and prescriptive roles. Prior work includes simulations calibrated to historical market data and those mimicking real investors. Some studies model trading teams or assign specialized roles to agents handling different data modalities before integrating recommendations, while others apply Retrieval-Augmented Generation (RAG) to ground responses in financial history. There are also efforts to simulate dynamic portfolio rebalancing. Despite advances, a key question remains: Do LLM agents truly resemble human investors in decision-making? Recent research comparing agent outputs with real-world analyses reveals discrepancies in structure, tone, and depth. LLM agents also show behavioral biases, such as conservatism, compared to professional institutions. Improving fidelity requires more than just data; agent design must incorporate richer representations of human intent, motivation, and context.

In summary, while LLM-based ABMs have advanced significantly, challenges remain in behavioral alignment, interpretability, and long-term strategic reasoning. Future work could explore hybrid models that integrate LLMs with domain-specific heuristics or fine-tune agents based on real investment logs.

2.2 Looking Ahead

Building on these insights, we propose focusing not only on prediction accuracy but also on how simulation systems can support human decision-making. Rather than simply retrieving information, next-generation decision-support systems should aim to generate actionable insights. Unlike traditional benchmarks that emphasize factual correctness, evaluations must account for how systems assist users in forming better judgments, beyond simple buy/sell recommendations. This includes the generation of detailed reasoning and forward-looking analysis. Future evaluations should also prioritize real user studies, measuring not only content quality but the impact on actual decision-making.

However, evaluating such systems presents challenges. Whether users behave similarly without real financial stakes and reproducibility across user groups are major concerns. Existing evaluation frameworks offer starting points but often overlook domain-specific variables like risk tolerance and financial literacy.

As LLM-based systems evolve from passive information providers to active decision-support agents, our evaluation methods must also evolve. By centering on real users and outcomes, we can better gauge how these systems enhance human expertise in high-stakes decision-making.

3 Recommender Systems for Investors

Recommender systems in the finance space aim to provide an easy and reliable means to identify effective investment options for a customer [21]. They exist as a low-cost solution to the significant challenges new entrants to the finance space experience, due to the complexity of financial markets; time needed to perform market research; difficulties in quantifying investment risks; as well as challenges in relating these to the personal circumstances of the customer [10]. Fundamentally, these systems are financial asset rankers, i.e. for a customer, they produce a ranking of financial assets to present, where assets deemed more suitable to the customer are ranked higher. The core ranking technology is then served by an investment platform, along with support tools, such as asset search and portfolio management [14].

However, how to effectively rank assets is a subject of much debate, and a wide variety of techniques have been proposed, leveraging a range of information sources, including investment transactions, pricing data, news and social networks, among others. Approaches can be divided broadly into unpersonalised vs. personalised approaches. Despite the intuitive need to personalise for the customer, the vast majority of financial recommender systems are unpersonalised [3, 5, 12, 18, 19]. The theory here is that regardless of the customer's circumstances, the goal is to maximise profit, and profit is not a function of the user but the market, hence personalisation is unnecessary. This has led to price-based or asset-based recommenders, which only consider asset-related information (e.g. prices and news) to suggest useful investments [17, 19]. The alternative is to attempt to personalise for the customer, typically using past investment transactions, either as a direct collaborative filtering signal, or by looking for patterns in prior investments [4, 7, 9, 11, 20]. The primary barriers to performing personalisation in this way is the lack of quality data to both drive and evaluate such systems [13]. Real investment data has commercial value and so is not freely available, moreover such data is complex and difficult to interpret as investors both make mistakes and do not record their intentions. Hence, to-date, personalised financial recommenders have been theoretically examined, but have seen little practical application.

On the other hand, personalisation for financial recommendation is starting to be re-examined. In particular, there is increased awareness that the value of human financial advisors stems from their ability to personalize investment guidance to clients' specific needs [6], rather than simply presenting suitable assets. Meanwhile, recent advancements in conversational AI and generative language models can facilitate human-like conversations with 'AI financial advisors'. It is these that we see as the next evolution of financial recommender systems, as they allow for information about each customer to be collected naturally through conversation, enabling intelligent personalisation without the need for noisy investment transaction data. Indeed, it has been demonstrated that an AI agent can provide equivalent performance to a human financial advisor when eliciting investment preferences from a customer, and that this can be used to guide a user to better select assets to invest in [16]. However, such agents are still in their infancy, with significant research and development needed to avoid preference hallucination and identify contradictions in conversations [15]. Furthermore, little work has been done integrating AI financial advisors with existing financial asset recommenders, and AI advisors currently lack the ability to reason about the recommendations that they produce, both rich areas for future research.

4 RAG for Investment Research Management Systems

Investment research workflows in hedge funds and asset management firms are traditionally hampered by manual processing of large volumes of unstructured documents such as 10-K filings, earnings calls, and broker reports. These workflows typically involve keyword searches, manual data extraction, and spreadsheet-based collation, leading to inefficiencies, inconsistencies, and delayed decision-making. To address these issues, Retrieval-Augmented Generation (RAG) systems have emerged as a transformative approach, integrating retrieval-based methods and generative language models (LLMs) to significantly improve the precision, relevance, and speed of financial information extraction [2].

Embedding quality lies at the heart of effective RAG [1]. However, embedding models in finance face distinct challenges, including noisy training data from loosely related web-crawled pairs and difficulties in calibrating the appropriate level of negative samples. Overly simple negatives provide limited learning, while excessively complex ones degrade model performance. To tackle these problems, we applied two targeted methodologies: task-specific data filtering, leveraging LLM-based similarity scoring for precise data curation; and synthetic data generation using expert-crafted few-shot prompts [1]. These strategies enabled training on over one billion tokens extracted from financial documents, fine-tuned with analyst-supervised feedback. Consequently, our embedding model attained first place in the HuggingFace Massive Text Embedding Benchmark (MTEB) as of March 2025, surpassing models from OpenAI and NVIDIA, with empirical tests showing approximately 17 percent improvement in retrieval accuracy, particularly notable in scenarios involving financial jargon and synonyms [1].

Recognizing that generic models inadequately capture financial terminology and relationships (e.g., differentiating "CapEx" and "Investments"), we further enhanced our embeddings through synthetic, domain-enriched data. This domain-specific approach significantly improved the accuracy and relevance of financial question answering and thematic searches [1].

To rigorously evaluate our system's real-world applicability, we developed FinDER (Financial Dataset for RAG), comprising 5,703 expert-generated query–evidence–answer triplets reflecting realistic analyst search behavior [2]. Unlike general-purpose QA sets, FinDER captures short, ambiguous, jargon-heavy queries, with over 90 percentages of queries containing three or more domain-specific terms. Evaluations highlighted the superiority of semantic retrieval over lexical methods, with semantic context boosting LLM accuracy by approximately 20 percentage points [2]. Nonetheless, a 30-percentage-point performance gap compared to ideal contexts remains, underscoring areas for further embedding and retrieval enhancement.

To streamline the extraction of structured key performance indicators (KPIs) from unstructured financial texts, we developed an LLM-native agentic extraction pipeline built upon Claude 3.7, Apache Spark orchestration, MLFlow monitoring, and Delta Lake storage—all integrated within Databricks and AWS infrastructures. This agent autonomously parses and structures data, drastically reducing the effort of manually compiling KPIs. Practical deployment has demonstrated precision levels exceeding 91.2 percent across over 100 financial metrics, translating to approximately 90 percent time savings compared to manual extraction. This marks the industry's first large-scale, production-grade PDF-to-SQL pipeline, now widely adopted by several top global hedge funds managing assets exceeding $100 billion.

Furthermore, recognizing the strategic value of thematic investing, we developed advanced thematic scoring methodologies using context-aware LLMs [8]. Traditional approaches relying on keyword counts and vector-based sentiment analysis are often insufficiently nuanced to capture strategic investment signals effectively. Our two-step thematic scoring method first identifies theme-relevant content within documents, then dynamically filters and evaluates this content for relevance and sentiment intensity. This approach robustly detects implicit themes—for example, associating "data centers" with terms like "server loads" or "cloud computing"—and quantifies firm sentiment toward these themes [8]. Empirical evaluations confirm that positive thematic sentiment scores correlate significantly with short-term stock outperformance, validating thematic scoring as a critical component of alpha generation.

Implemented in practice, our system evaluates thematic exposure and sentiment for over 1,000 companies weekly across more than 100 themes, providing portfolio managers and quantitative analysts a precise, automated, and reliable analytical tool for strategic decision-making and risk assessment [8].

In summary, our integration of sophisticated embedding enhancements [1], agent-based structured data extraction, domain-specific evaluation via FinDER [2], and granular thematic scoring [8] represents a comprehensive, industry-proven RAG system. Its adoption across leading hedge funds demonstrates tangible benefits including workflow acceleration, deeper analytical insights, and measurable alpha generation, underscoring the value and transformative potential of specialized retrieval and generative AI within investment research.

5 Using LLMs for Portfolio Selection

Large Language Models (LLMs), with their ability to simultaneously analyze vast amounts of news and documents as well as perform complex reasoning, hold significant promise for financial investment applications. However, despite substantial progress, concerns about hallucination have not been entirely eliminated. In this context, directly entrusting LLMs with investment decision-making remains challenging. This tutorial aims to explore how LLMs can be effectively utilized in investment decision processes under such uncertainty.

Specifically, the tutorial will cover three key subtopics:

Fine-tuning LLM Embedding Models for Portfolio Selection: In many investment-related tasks that leverage text data, it is common to embed textual information and use it as input to ML/AI models. In this tutorial, we introduce an approach to fine-tune embeddings such that investment ideas extracted from text are aligned with the embeddings of relevant stocks or ETFs. By ensuring that stocks and ETFs that share similar investment themes are located closer in the embedding space, we enable investors to explore investment opportunities more freely and intuitively based on natural language descriptions.

Understanding Biases of LLMs in Investment Contexts: Although biases in LLMs, such as those related to political ideology, gender, and race, are well documented, their biases in financial investment contexts remain underexplored. We examine whether various LLMs exhibit tendencies toward aggressive or conservative investment strategies, preferences for certain sectors or themes, or inclinations toward momentum or contrarian approaches in market timing. This section will provide a systematic investigation into the types of bias that LLMs can introduce into investment decision-making.

Incorporating LLM Outputs in Portfolio Optimization: Finally, we demonstrate how to incorporate LLM-generated views, such as opinions about specific stocks or sectors—into portfolio optimization using the Black-Litterman model. Originally developed to integrate subjective views into mean-variance optimization, the Black-Litterman framework provides a natural way to blend LLM outputs into systematic portfolio construction. We will showcase how different LLM models and variations in prompting strategies can affect the resulting portfolios.

Through these three components, this tutorial will provide a comprehensive overview of the opportunities and challenges of integrating LLMs into portfolio selection and optimization.

6 Conclusion

Financial information retrieval is undergoing a paradigm shift driven by advances in large language models and retrieval-augmented generation systems. Our tutorial illustrates how these technologies are reshaping investment research, decision support, and portfolio management. However, challenges such as behavioral alignment, personalization, and system evaluation remain open. Continued interdisciplinary efforts are needed to fully realize the potential of intelligent, retrieval-augmented financial decision systems.

References

[1] Chanyeol Choi, Junseong Kim, Seolhwa Lee, Jihoon Kwon, Sangmo Gu, Yejin Kim, Minkyung Cho, and Jy yong Sohn. 2024. Linq-Embed-Mistral Technical Report. https://doi.org/10.48550/arXiv.2412.03223 arXiv:2412.03223 [cs.CL] Accessed: 2024-12-04.

[2] Chanyeol Choi, Jihoon Kwon, Jaeseon Ha, Hojun Choi, Chaewoon Kim, Yongjae Lee, Jy yong Sohn, and Alejandro Lopez-Lira. 2025. FinDER: Financial Dataset for Question Answering and Evaluating Retrieval-Augmented Generation. https://doi.org/10.48550/arXiv.2504.15800 arXiv:2504.15800 [cs.IR] ICLR 2025 Workshop Advances in Financial AI, Accessed: 2025-04-23.

[3] Shibo Feng, Chen Xu, Yu Zuo, Guo Chen, Fan Lin, and Jianbing XiaHou. 2022. Relation-aware dynamic attributed graph attention network for stocks recommendation. *Pattern Recognition* 121 (2022), 108119:1–108119:12. https://doi.org/10.1016/j.patcog.2021.108119

[4] Reyes Michaela Denise Gonzales and Carol Anne Hargreaves. 2022. How can we use artificial intelligence for stock recommendation and risk management? A proposed decision support system. *International Journal of Information Management Data Insights* 2, 2 (2022), 100130:1–100130:10. https://doi.org/10.1016/j.jjimei.2022.100130

[5] Chien-Feng Huang. 2012. A hybrid stock selection model using genetic algorithms and support vector regression. *Applied Soft Computing* 12, 2 (2012), 807–818. https://doi.org/10.1016/j.asoc.2011.10.009

[6] Francis M. Kinniry Jr., Colleen M. Jaconetti., Michael A. DiJoseph., Yan Zilbering., Donald G. Bennyhoff., and Georgina Yarwood. 2020. *Putting a value on your value: Quantifying Vanguard Adviser's Alpha in the UK.* Technical Report. The Vanguard Group, Valley Forge, Pennsylvania, USA.

[7] Youngbin Lee, Yejin Kim, Javier Sanz-Cruzado, Richard Mccreadie, and Yongjae Lee. 2024. Stock Recommendations for Individual Investors: A Temporal Graph Network Approach with Mean-Variance Efficient Sampling. In *Proceedings of the 5th ACM International Conference on AI in Finance*. 795–803.

[8] Alejandro Lopez-Lira, Chanyeol Choi, Yoon Kim, Jihoon Kwon, and Suyeol Yun. 2025. Thematic Scoring: Quantifying Contextual Narratives using Language Models. *SSRN Electronic Journal* (2025). https://doi.org/10.2139/ssrn.5220998 Accessed: 2025-04-15.

[9] Johannes Luef, Christian Ohrfandl, Dimitris Sacharidis, and Hannes Werthner. 2020. A Recommender System for Investing in Early-Stage Enterprises. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing (SAC 2020)*. Association for Computing Machinery, Online, 1453–1460. https://doi.org/10.1145/3341105.3375767

[10] Richard McCreadie, Konstantinos Perakis, Maanasa Srikrishna, Nikolaos Droukas, Stamatis Pitsios, Georgia Prokopaki, Eleni Perdikouri, Craig Macdonald, and Iadh Ounis. 2022. Next-generation personalized investment recommendations. *Big Data and Artificial Intelligence in Digital Finance: Increasing Personalization and Trust in Digital Finance using Big Data and AI* (2022), 171–198.

[11] Cataldo Musto, Giovanni Semeraro, Pasquale Lops, Marco de Gemmis, and Georgios Lekkas. 2015. Personalized finance advisory through case-based recommender systems and diversification strategies. *Decision Support Systems* 77 (2015), 100–111. https://doi.org/10.1016/j.dss.2015.06.001

[12] Tong-Seng Quah and Bobby Srinivasan. 1999. Improving returns on stock investment through neural network selection. *Expert Systems with Applications* 17, 4 (1999), 295–301. https://doi.org/10.1016/S0957-4174(99)00041-X

[13] Javier Sanz-Cruzado, Nikolaos Droukas, and Richard McCreadie. 2024. FAR-Trans: An Investment Dataset for Financial Asset Recommendation. In *Proceedings of the IJCAI-24 Workshop on Recommender Systems in Finance (Fin-RecSys).*

[14] Javier Sanz-Cruzado, Edward Richards, and Richard McCreadie. 2024. FAR-AI: A Modular Platform for InvestmentR ecommendation in the Financial Domain. In *Proceedings of the 46th European Conference on Information Retrieval, (ECIR 2024), Part V.* Glasgow, United Kingdom, 267–271. https://doi.org/10.1007/978-3-031-56069-9_30

[15] Takehiro Takayanagi, Kiyoshi Izumi, Javier Sanz-Cruzado, Richard Mccreadie, and Iadh Ounis. 2025. Are Generative AI Agents Effective Personalized Financial Advisors?. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval.*

[16] Takehiro Takayanagi, Masahiro Suzuki, Kiyoshi Izumi, Javier Sanz-Cruzado, Richard Mccreadie, and Iadh Ounis. 2025. FinPersona: An LLM-Driven Conversational Agent for Personalized Financial Advising. In *Proceedings of the 47th European Conference on Information Retrieval (ECIR 2025).* Lucca, Italy.

[17] Wenting Tu, Min Yang, David W. Cheung, and Nikos Mamoulis. 2018. Investment recommendation by discovering high-quality opinions in investor based social networks. *Information Systems* 78 (2018), 189–198. https://doi.org/10.1016/j.is.2018.02.011

[18] Mei-Chen Wu, Szu-Hao Huang, and An-Pin Chen. 2024. Momentum portfolio selection based on learning-to-rank algorithms with heterogeneous knowledge graphs. *Applied Intelligence* 54, 5 (2024), 4189–4209. https://doi.org/10.1007/S10489-024-05377-2

[19] Hongyang Yang, Xiao-Yang Liu, and Qingwei Wu. 2018. A Practical Machine Learning Approach for Dynamic Stock Recommendation. In *Proceedings of the 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE 2018).* IEEE, New York, NY, USA, 1693–1697. https://doi.org/10.1109/TrustCom/BigDataSE.2018.00253

[20] Xiaoxue Zhao, Weinan Zhang, and Jun Wang. 2015. Risk-Hedged Venture Capital Investment Recommendation. In *Proceedings of the 9th ACM Conference on Recommender Systems (RecSys 2015).* Association for Computing Machinery, Vienna, Austria, 75–82. https://doi.org/10.1145/2792838.2800181

[21] Dávid Zibriczky. 2016. Recommender systems meet finance: a literature review. In *Proceedings of the 2nd International Workshop on Personalization & Recommender Systems in Financial Services (FinRec 2016).* 1–10.