A Survey of Large Language Models Agents Market Microstructure Algorithmic Trading Natural Language Processing and Financial Forecasting

www.surveyx.cn

Abstract

In the multidisciplinary domain of financial markets, the integration of Large Language Models (LLMs), autonomous agents, market microstructure, algorithmic trading, and natural language processing (NLP) offers transformative potential for enhancing market analysis and decision-making processes. This survey explores the synergistic application of these technologies, highlighting their capabilities in processing unstructured financial data, optimizing trading strategies, and predicting market trends. LLMs, with their proficiency in generating human-like text, facilitate the extraction of actionable insights, while their integration with multimodal data sources enhances analytical depth. Autonomous agents, leveraging LLMs, improve decision-making efficacy in trading environments by simulating human-like reasoning and adaptability. The study of market microstructure, coupled with algorithmic trading, provides insights into price formation and liquidity dynamics, optimizing trade execution. Despite their transformative potential, challenges persist, including computational resource demands, data privacy concerns, and model interpretability issues. Addressing these challenges through interdisciplinary collaboration and ethical frameworks is crucial for responsible deployment. Future research directions emphasize the need for lightweight models, enhanced bias mitigation, and improved interpretability to fully harness the capabilities of LLMs in financial markets. This comprehensive survey underscores the importance of ongoing innovation and ethical considerations in leveraging advanced AI technologies to navigate the complexities of modern financial environments.

1 Introduction

1.1 Multidisciplinary Domain Overview

The multidisciplinary domain integrating Large Language Models (LLMs), agents, market microstructure, algorithmic trading, natural language processing (NLP), and financial forecasting represents a convergence of advanced technologies aimed at enhancing the understanding and functionality of financial markets. LLMs have revolutionized data augmentation strategies in NLP, presenting new opportunities and addressing existing challenges, as noted by Ding et al. [1]. Despite their impressive performance in various NLP tasks, challenges persist in domain-specific applications requiring advanced analytical capabilities, as highlighted by Kang et al. [2].

The integration of multimodal large language models (MLLMs) is crucial, as they adeptly process and synthesize diverse data types—text, images, audio, and video—thereby enhancing financial data analysis. Jin et al. [3] emphasize the architectural innovations and training methodologies that drive these models, while Han et al. [4] explore advancements in multimodal pretrained models that foster innovation across sectors, including finance.

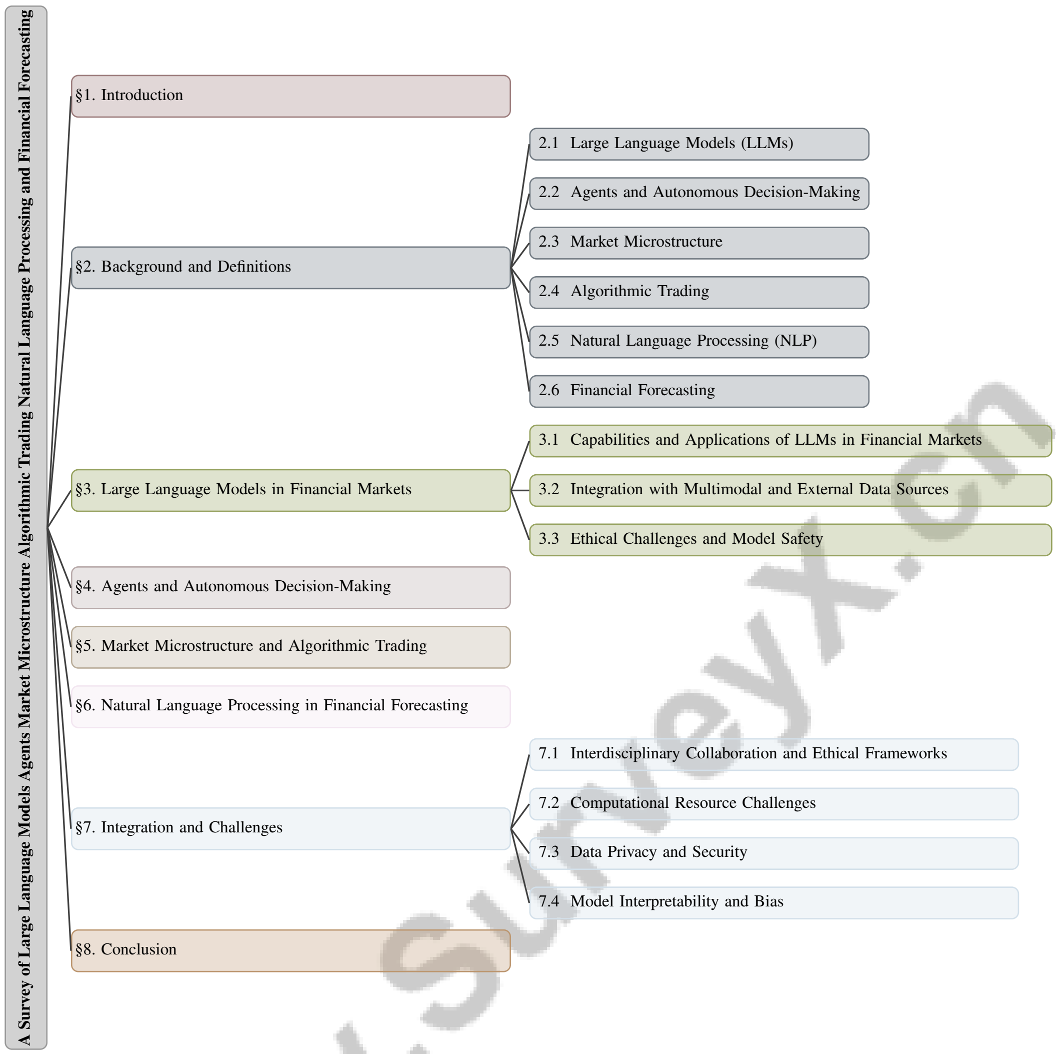


Figure 1: chapter structure

Autonomous agents are vital in automating and optimizing decision-making processes within financial markets. Li [5] discusses the development of automated workflows for LLM-based AI agents, which streamline complex manual designs, thus enhancing trading system efficiency. Furthermore, the exploration of self-identification in intelligent agents by Li [6] underscores the importance of computational psychoanalysis in improving AI functionality and adaptability.

Market microstructure and algorithmic trading are essential for understanding trading dynamics and executing trades optimally. Advanced NLP techniques enhance these components by enabling the extraction of actionable insights from complex financial texts, such as earnings calls and reports, and facilitating accurate market trend predictions through methods like Sequential Knowledge-Guided Prompting and multimodal feature fusion [7, 8, 9, 10].

Financial forecasting leverages these technologies to predict market movements and guide strategic decisions. The intelligence embedded in 6G networks, as explored by Long et al. [11], enhances cognitive capabilities and security, fostering innovation in financial applications.

The challenges and opportunities arising from these advancements, particularly concerning trust and ethical considerations in AI automation, necessitate continuous examination of the methodologies involved. This multidisciplinary approach not only enhances the efficiency of financial market operations but also requires ongoing scrutiny to address emerging challenges and knowledge gaps, as discussed by Zhou et al. [12].

AI-generated, for reference only.

1.2 Significance of Each Component

Each component in this survey significantly advances financial market capabilities through the integration of cutting-edge technologies. LLMs enhance reasoning performance across various tasks, particularly in complex scenarios, exemplified by Chain of Thought (CoT) techniques [13]. They simplify deployment by reducing the need for specialized models, as demonstrated by the multi-modal, multi-task agent Gato [14]. The LLMs4OL benchmark further highlights their importance in ontology learning tasks, crucial for structuring and understanding financial data [15].

Autonomous agents automate decision-making in financial markets, addressing complexities and time constraints associated with manual processes [5]. These agents enhance operational capabilities by optimizing actions based on a comprehensive understanding of utility rather than solely improving prediction accuracy [16]. Moreover, the integration of LLMs within 6G networks creates an intelligent network architecture that optimizes performance and operational capabilities [11].

Market microstructure and algorithmic trading are vital for understanding trading dynamics and executing trades optimally. Their significance is further accentuated by advancements in NLP, which facilitate the extraction of actionable insights from financial texts, enhancing financial forecasting accuracy. Accurate decision-making often relies on uncertain numerical data, which can be challenging to interpret [17].

Financial forecasting utilizes these advanced technologies to predict market movements and inform strategic decisions, ensuring markets remain agile and responsive to emerging trends. A comprehensive survey of LLM-related concepts, including architectural innovations and training strategies, emphasizes the importance of these components in enhancing financial market operations' efficiency and effectiveness [18].

1.3 Interconnections and Synergies

The interconnections and synergies among LLMs, agents, market microstructure, algorithmic trading, NLP, and financial forecasting create a complex framework that enhances financial market functionality and efficiency. LLMs provide a foundational layer that supports actionable insights extraction, crucial for informed decision-making in financial contexts [1]. When integrated with autonomous agents, these insights enhance agents' capabilities to make data-driven decisions, optimizing trading strategies and improving market outcomes [5].

The synergy between market microstructure and algorithmic trading is evident as algorithmic systems utilize microstructural data to execute trades under optimal conditions, influencing market liquidity and price discovery. This integration is augmented by NLP techniques, facilitating the interpretation of financial news and reports, which provides nuanced insights into market sentiments and trends [17]. Consequently, predictive models can incorporate both quantitative data and qualitative insights derived from textual analysis, leading to more accurate financial forecasting.

Furthermore, incorporating LLMs within 6G networks exemplifies the convergence of advanced AI technologies and communication infrastructures, creating intelligent networks that enhance operational capabilities and ensure secure financial transactions [11]. This interconnected framework not only improves financial market operations but also fosters innovation by enabling the development of new applications and services that leverage the combined strengths of these technologies [3].

1.4 Structure of the Survey

This survey is structured to provide a comprehensive exploration of the multidisciplinary domain integrating LLMs, agents, market microstructure, algorithmic trading, NLP, and financial forecasting. It begins with an introduction that establishes the significance and interconnections of these components within financial markets. Following this, the background and definitions section contextualizes each core concept, laying a foundation for subsequent discussions.

The survey then examines the role of LLMs in financial markets, highlighting their capabilities in processing financial data and contributing to market analysis and decision-making. This is followed by an in-depth exploration of various types of agents involved in autonomous decision-making, emphasizing their integration within trading systems and the resulting effects on market dynamics, including price formation, volatility, and order flow persistence, along with implications for market stability and regulatory considerations. The analysis also investigates interactions among AI traders, fundamental traders, and noise traders within a multi-agent framework, supported by empirical data and simulations [19, 20, 21, 22].

Subsequent sections delve into market microstructure and algorithmic trading, analyzing technological innovations that enhance trading systems and their influence on market processes. The role of NLP in financial forecasting is then investigated, emphasizing techniques for processing financial textual data and recent advancements that improve predictive accuracy.

The survey culminates in a discussion on the integration and challenges associated with these technologies in financial markets, addressing interdisciplinary collaboration, ethical frameworks, computational resource challenges, data privacy, security, and model interpretability. The conclusion synthesizes key findings, emphasizing the effectiveness of novel prompting strategies—such as Role-Playing (RP) and Chain-of-Thought (CoT)—in enhancing LLM performance in sentiment analysis, as well as challenges encountered in multimodal long-form summarization, particularly concerning financial reports. It also highlights critical areas for future research, including the exploration of LLM capabilities in handling complex datasets and the need to address existing knowledge gaps, reinforcing the necessity for ongoing investigation in this rapidly evolving field [8, 23].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Large Language Models (LLMs)

Large Language Models (LLMs) represent a significant advancement in AI, crucial for financial applications due to their ability to comprehend and generate human-like text. Their deployment is challenged by their computational intensity [24]. LLMs excel in processing extensive textual data, thereby enhancing decision-making and strategic planning in financial markets [9]. They automate the analysis of unstructured data, vital for ontology learning, which involves term identification, taxonomy construction, and semantic relationship extraction [12]. This is essential for structuring complex financial datasets and enhancing market analysis frameworks. The integration of multimodal data further enhances LLMs' analytical capabilities, providing deeper insights into financial trends [4].

LLMs improve reasoning, particularly in complex tasks involving graph structures and multimodal data, although distribution shifts can affect performance in multimodal reasoning and Chain-of-Thought processes [25]. Innovative training methods, such as pretraining and supervised fine-tuning, are crucial for enhancing LLMs' reasoning abilities in financial applications. LLMs simulate complex market interactions, addressing the challenge of replicating dynamic human trading behaviors [26]. Their ability to enhance clustering through improved semantic representations aids in event detection [27]. Lifelong learning capabilities enable LLMs to adapt to evolving data and user preferences, addressing issues like catastrophic forgetting [28]. However, understanding complex interactions and mitigating selection bias remain challenges impacting LLM reliability [29].

In predicting stock movements, LLMs analyze textual data, such as news articles influencing prices [9]. Their predictive power is enhanced through structured prompts, incorporating role-playing and logical reasoning elements [23]. The rise of generative AI tools based on LLMs raises privacy concerns, as models may inadvertently extract sensitive information during training [30]. Integrating LLMs with optimization algorithms is vital for managing operations in financial markets, especially where traditional derivative-based methods fall short [31]. This synergy enhances LLMs' practical applications, addressing computational challenges and adapting to increasing market complexity [11]. Combining LLMs with Knowledge Graphs (KGs) presents opportunities for innovation, leveraging both technologies' strengths.

LLMs underpin enhanced analytical capabilities in financial markets, necessitating ongoing research to address deployment challenges, including ethical and legal considerations [32]. Their ability to generate complex instructions efficiently is crucial for aligning LLM training with user expectations and instruction tuning.

2.2 Agents and Autonomous Decision-Making

Agents, as autonomous decision-making entities, are crucial in financial markets for emulating human-like decision processes in complex environments. They operate independently, analyzing market data to optimize trading strategies and enhance market efficiency [33]. The integration of advanced AI technologies, particularly LLMs, significantly boosts agents' adaptability and performance, enabling complex tasks like code generation and data analysis beyond traditional systems' capabilities [34]. However, the rigidity of LLM agents, constrained to predefined actions, limits their adaptability in dynamic environments [35]. This necessitates developing flexible systems capable of adjusting actions in response to market changes. Additionally, enabling intelligent agents to achieve self-identification by simulating human psychological processes poses a significant challenge [6].

Multi-agent systems, employing a collaborative approach, are vital for modeling market dynamics and understanding AI traders' impacts on price formation and volatility [21]. These systems enhance efficiency by allowing agents to communicate and coordinate actions, although LLM-based multi-agent systems often encounter communication inconsistencies that impede their effectiveness in complex tasks [36]. The emergence of generalist agents, such as Gato, exemplifies the potential of integrating diverse modalities and task types into a single model while maintaining performance [14]. These agents can perform various tasks, from trading to data analysis, streamlining operations within financial markets. Nonetheless, constructing autonomous agents that leverage LLMs to replicate human-like decision-making in complex, uncertain environments remains a challenge [37].

2.3 Market Microstructure

Market microstructure involves analyzing the processes governing security trading, focusing on price formation in response to supply and demand fluctuations. It examines liquidity dynamics and trading costs, revealing that market liquidity is often low, necessitating incremental execution of large orders over time, which leads to persistent long-memory characteristics in order flow. The field combines theoretical and empirical insights to predict market phenomena, including market impact, bid-ask spreads, order book dynamics, and volatility, while assessing AI traders' behaviors on market stability and regulatory frameworks [19, 21]. It explores how information is integrated into prices, the role of market makers, and the effects of trading rules and regulations on market outcomes.

Crypto assets present unique challenges for market microstructure analysis. Unlike traditional equity markets, crypto markets exhibit different liquidity dynamics and price formation processes due to their decentralized nature and absence of central authorities, necessitating specialized models and strategies for understanding market behavior [20]. A significant challenge in market microstructure research is the scarcity of comprehensive data sources, limiting thorough analyses and model development. The importance of examining licensing terms and the difficulties in extracting value from alternative data sources are emphasized [38]. These challenges highlight the need for innovative data collection and analysis approaches to enhance market microstructure understanding.

The slow digestion of supply and demand fluctuations is a core issue in market microstructure studies. Bouchaud [19] discusses how these fluctuations influence price formation and liquidity dynamics, impacting overall market efficiency. Understanding these processes is critical for market participants, enabling strategies that optimize trading outcomes and mitigate risks associated with volatility.

2.4 Algorithmic Trading

Algorithmic trading uses computer algorithms to automate trade execution in financial markets. These algorithms make decisions regarding timing, price, and quantity of trades based on predefined criteria, allowing rapid order execution and optimization of trading strategies. The primary advantage of algorithmic trading is its ability to process vast amounts of market data and execute trades at speeds beyond human capabilities, enhancing market liquidity and improving price formation mechanisms through incremental trading of large orders [19, 21, 9, 10].

Managing noisy and imperfect data is critical for accurate trading decisions. The Noisy Mixture of Experts (Noisy MoE) method addresses this by using a least trimmed squares algorithm to refine parameter estimation for experts and the gating function [39], enhancing trading algorithm robustness in the dynamic financial landscape. Integrating multimodal large language models (MLLMs) into algorithmic trading systems broadens their capabilities by facilitating the analysis of diverse data forms, including text, images, and audio. However, challenges remain in pretraining and fine-tuning these models for varied modalities and managing high computational costs [4]. Despite these hurdles, MLLMs in algorithmic trading hold promise for improving strategy accuracy and efficiency through a more comprehensive understanding of market conditions.

2.5 Natural Language Processing (NLP)

Natural Language Processing (NLP) is crucial in finance, enabling the extraction and interpretation of insights from extensive unstructured textual data. This capability is essential for analyzing financial reports, news articles, and other text-based information affecting market dynamics. The integration of LLMs has significantly advanced NLP's ability to handle complex financial data, processing and generating human-like text with high accuracy. LLMs are utilized in data augmentation tasks, employing strategies for data creation, labeling, and reformation to enhance model performance [40].

Multimodal models that combine text with other data forms, such as images and audio, are essential for comprehensive financial analysis. Evaluations of datasets and benchmarks for multimodal model assessment ensure that NLP applications can effectively meet the diverse needs of financial markets [41]. However, challenges such as the 'black box' nature of LLMs, which obscures decision-making processes, pose significant concerns [41]. Additionally, LLMs' propensity to generate factually incorrect responses, known as hallucinations, can mislead users and hinder decision-making [42].

The financial sector requires specialized NLP models due to the complexity of financial data and tasks. A benchmark by Gkatzia et al. [17] evaluates the effectiveness of various information presentation strategies for uncertain data, contrasting Natural Language Generation (NLG) with traditional graphical methods, underscoring NLP's role in financial information presentation. Furthermore, deploying LLMs in financial applications must address privacy, bias, and fairness issues. Privacy-preserving mechanisms are essential to protect user privacy during data curation and training [30]. Additionally, LLMs may inadvertently learn and perpetuate harmful social biases, necessitating robust bias evaluation and mitigation techniques [43].

2.6 Financial Forecasting

Financial forecasting is essential for anticipating future market trends and outcomes, crucial for strategic decision-making and optimizing investment strategies. The integration of advanced technologies, such as LLMs, has significantly improved the accuracy and efficiency of financial forecasting. LLMs, particularly decoder-only models, have shown substantial utility in summarization and program repair tasks, vital for processing financial data [44]. Reliance on real-world data for forecasting presents challenges, especially in specialized domains with scarce or inaccessible data. Generating synthetic data has emerged as a vital solution for enhancing predictive model accuracy [45]. Moreover, integrating LLMs with optimization algorithms has been shown to boost performance across various applications, enhancing financial forecasting capabilities [31].

Benchmarking efforts, such as those evaluating BloombergGPT, emphasize the importance of assessing model performance on financial NLP tasks. These benchmarks highlight the effectiveness of domain-specific models compared to general-purpose ones, underscoring the need for specialized tools in financial forecasting [7]. Furthermore, deducing initial strategy distributions of agents in mix-game models is crucial for refining market predictions, enhancing the understanding of market dynamics and agent behavior [22].

3 Large Language Models in Financial Markets

In recent years, the role of Large Language Models (LLMs) in financial markets has garnered significant attention due to their potential to revolutionize various analytical processes. As we delve into the specific capabilities and applications of these models, it becomes evident that their integration into financial systems not only enhances data processing but also supports strategic decision-making. Figure 2 illustrates the hierarchical structure of LLMs in financial markets, showcasing their capabilities and applications, as well as their integration with multimodal and external data sources. This figure also addresses ethical challenges and model safety concerns, highlighting how LLMs enhance market analysis, support strategic decision-making, and provide comprehensive data processing solutions. It emphasizes the importance of addressing biases, ensuring model safety, and maintaining data privacy.

The following subsection will explore the multifaceted capabilities of LLMs, further highlighting their applications in financial contexts and the transformative impact they have on market analysis and business operations.

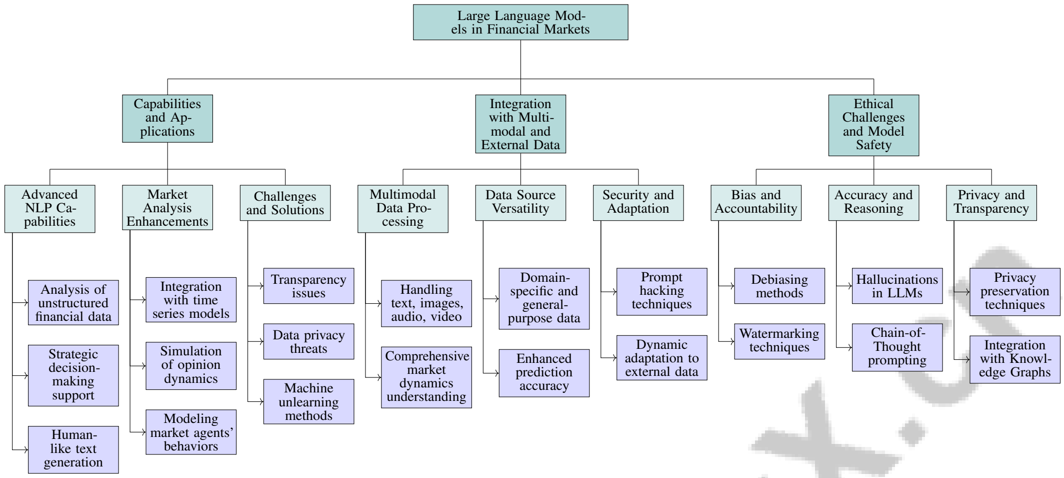


Figure 2: This figure illustrates the hierarchical structure of Large Language Models (LLMs) in financial markets, showcasing their capabilities and applications, integration with multimodal and external data sources, and addressing ethical challenges and model safety concerns. It highlights how LLMs enhance market analysis, support strategic decision-making, and provide comprehensive data processing solutions while emphasizing the importance of addressing biases, ensuring model safety, and maintaining data privacy.

3.1 Capabilities and Applications of LLMs in Financial Markets

Large Language Models (LLMs) have emerged as transformative tools in financial markets, offering advanced capabilities in natural language processing that facilitate the analysis of unstructured financial data and support strategic decision-making. These models excel in generating human-like text, enabling the interpretation of complex language constructs and providing insights essential for market analysis [46]. The integration of LLMs into financial systems has the potential to reshape business models by aligning technological investments with business objectives, creating new revenue streams, and enhancing operational efficiency. Frameworks like AutoFlow automate workflow generation for AI agents, streamlining the incorporation of LLMs into financial applications [40].

LLMs have demonstrated state-of-the-art performance on financial NLP tasks. Specialized models such as BloombergGPT enhance prediction accuracy by integrating with time series models, thereby improving forecasting and strategic planning capabilities [9]. The Chain of Thought (CoT) method, which automates reasoning chains, further enhances reasoning performance in financial market analysis [47]. Moreover, LLMs facilitate the simulation of opinion dynamics within networks, capturing the complexity of human interactions and market sentiment evolution, crucial for informed financial decision-making [25].

The application of LLMs extends to modeling market agents whose behaviors converge toward competitive equilibria, as demonstrated in studies evaluating their innovative application in simulating market dynamics [26]. Additionally, LLMs enhance the clustering process and event detection through keyword extraction, text embedding, summarization, and topic labeling [27]. These capabilities are critical for identifying trends and anomalies in financial data, thereby supporting proactive market strategies.

Despite their transformative potential, LLMs face challenges such as transparency issues, which can lead to misuse and overreliance due to stakeholders' limited understanding of these models' behaviors and limitations [48]. The risk of training data memorization also poses significant threats to data privacy, necessitating effective mitigation strategies. Machine unlearning methods offer a solution by selectively removing harmful knowledge while preserving valuable information [24].

The CREATOR framework enhances the problem-solving capabilities of LLMs by enabling them to create and modify tools based on specific problems, increasing their flexibility and adaptability in financial applications [49]. Furthermore, the integration of LLMs with Knowledge Graphs (KGs) through frameworks like KG-enhanced LLMs, LLM-augmented KGs, and Synergized LLMs + KGs offers opportunities for innovation, leveraging their respective strengths to overcome the limitations of both technologies [42].

As shown in Figure 3, this figure illustrates the capabilities and applications of Large Language Models (LLMs) in financial markets, focusing on advanced NLP tasks, market dynamics, and associated challenges and innovations. It highlights key tools and methods such as BloombergGPT for prediction accuracy, the Chain of Thought method for reasoning enhancement, and the CREATOR framework for problem-solving flexibility. This example underscores the transformative potential of LLMs in financial analysis and decision-making processes. The first subfigure, "Identification, Screening, Eligibility, and Included Records," outlines a systematic approach to research, showcasing the progression through critical stages such as identification and screening of records, pivotal for ensuring the reliability and relevance of data used in financial analyses. Meanwhile, the second subfigure, "Focus Category Focus Item Questions," highlights the structured categorization of financial inquiries, where LLMs can be instrumental in parsing and responding to complex questions across various domains like financial indicators, managerial roles, and future market outlooks. Together, these visual examples underscore the potential of LLMs to enhance decision-making processes by streamlining data management and providing insightful analyses tailored to specific financial contexts [50, 10].

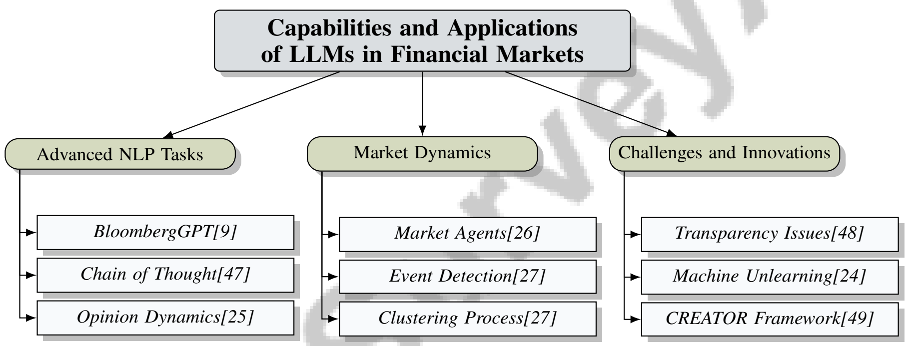


Figure 3: This figure illustrates the capabilities and applications of Large Language Models (LLMs) in financial markets, focusing on advanced NLP tasks, market dynamics, and associated challenges and innovations. It highlights key tools and methods such as BloombergGPT for prediction accuracy, the Chain of Thought method for reasoning enhancement, and the CREATOR framework for problem-solving flexibility. The figure underscores the transformative potential of LLMs in financial analysis and decision-making processes.

3.2 Integration with Multimodal and External Data Sources

The integration of Large Language Models (LLMs) with multimodal and external data sources represents a significant advancement in enhancing analytical capabilities within financial markets. This integration allows for the processing and interpretation of diverse data forms, such as text, images, audio, and video, thereby providing a more comprehensive understanding of market dynamics and trends. The development of models like MoDeGPT, which employs modular optimization techniques to compress multiple matrices within the functional modules of a Transformer, exemplifies the potential of LLMs to efficiently handle complex multimodal data [51].

The ability of LLMs to leverage both domain-specific and general-purpose data, as demonstrated by BloombergGPT, underscores their versatility in performing financial tasks while maintaining general NLP capabilities [7]. This dual capability is crucial for financial applications, where the integration of diverse data sources can enhance prediction accuracy and strategic decision-making. Furthermore, the structured prompting approach employed by FLAME enhances the comprehension of hypernymy relationships in taxonomies, facilitating more nuanced data analysis and interpretation [52].

The MULTI framework introduces a variety of question types, including multiple-choice, fill-in-the-blank, and open-ended questions, which enhance the capabilities of LLMs in understanding and

AI-generated, for reference only.

processing multimodal data within financial markets [53]. This diversity in question types allows LLMs to adapt to different analytical scenarios, providing more robust and flexible solutions for financial analysis.

In addition to multimodal integration, the dynamic adaptation of scene descriptions and procedural generation in the multi-agent framework of 3D-GPT highlights the potential for LLMs to interact with and adapt to external data sources in real-time [54]. This capability is particularly beneficial in financial markets, where timely and accurate data interpretation is essential for informed decision-making.

Moreover, the identification and categorization of various prompt hacking techniques, as explored in recent research, enhance the understanding of the security landscape surrounding LLMs, ensuring that these models can be deployed safely and effectively in financial applications [55]. By addressing security concerns and optimizing data integration processes, LLMs can provide more reliable and comprehensive insights into financial markets.

3.3 Ethical Challenges and Model Safety

The deployment of Large Language Models (LLMs) in financial markets introduces significant ethical challenges and model safety concerns that demand comprehensive strategies for mitigation. As illustrated in Figure 4, which depicts the hierarchical structure of these challenges, key areas of focus include bias, hallucinations, model safety, and knowledge integration. One of the primary ethical issues is the presence of biases within LLMs, which can result in skewed decision-making processes and potentially discriminatory outcomes. To address this, robust debiasing methods are necessary to ensure fair treatment across diverse scenarios [29]. Additionally, the risk of LLMs generating biased or harmful content highlights the importance of implementing effective watermarking techniques to mitigate misuse and enhance accountability [56].

Another critical challenge is the phenomenon of hallucinations, where LLMs produce factually incorrect or misleading information. This issue is particularly concerning in financial markets, where accuracy is crucial. Despite advancements, LLMs continue to struggle with complex reasoning tasks, suggesting that existing benchmarks may not fully capture their reasoning abilities [47]. Methodologies such as Chain-of-Thought (CoT) prompting have been developed to elicit complex reasoning in large models, improving performance on challenging tasks [57]. However, the complexities of knowledge integration often introduce noise and require additional computational resources, complicating the reasoning process [42].

Model safety is also threatened by potential prompt hacking attacks, which can undermine the integrity of LLMs. These attacks are categorized into jailbreaking, injection, and leaking, each posing unique risks to model security. Ensuring model safety necessitates the development of robust frameworks that incorporate dynamic memory updates and real-time learning, enabling LLMs to adapt to evolving threats [58]. The integration of advanced frameworks offers a promising approach to enhancing the efficiency of fine-tuning LLMs while ensuring compliance with privacy regulations [32].

Transparency is another crucial aspect of model safety, with a human-centered approach emphasizing the importance of understanding the needs and contexts of various stakeholders when designing transparency mechanisms [48]. Enhancements in interpretability, such as the integration of Knowledge Graphs (KGs) with LLMs, have been shown to improve factual accuracy and user trust [59]. Moreover, the ability to process multimodal data effectively enhances interpretability and prediction accuracy, which is essential in financial applications [41].

Privacy protection remains a critical concern, with significant deficiencies identified in current LLMs' privacy safeguards. The benchmark highlights the importance of privacy preservation in LLMs and provides a structured approach to evaluate and improve privacy-preserving techniques [30]. These efforts underscore the necessity of safeguarding sensitive information, particularly in financial applications where data privacy is of utmost importance.

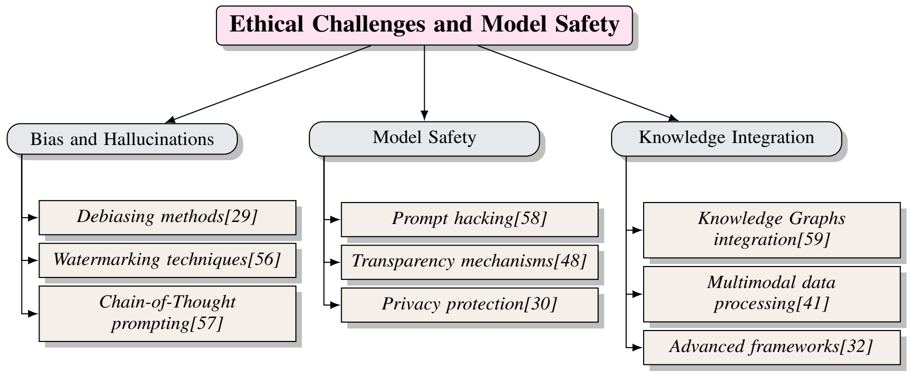


Figure 4: This figure illustrates the hierarchical structure of ethical challenges and model safety in large language model deployment, focusing on bias, hallucinations, model safety, and knowledge integration.

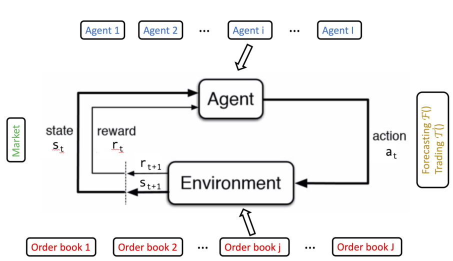
4 Agents and Autonomous Decision-Making

The development of AI agents and their frameworks is pivotal in enhancing operational efficiency within trading environments, optimizing decision-making processes, and improving market dynamics. This section reviews the taxonomy and frameworks crafted for AI agents in trading, highlighting methodologies that are essential for creating sophisticated agents capable of navigating the complexities of financial markets.

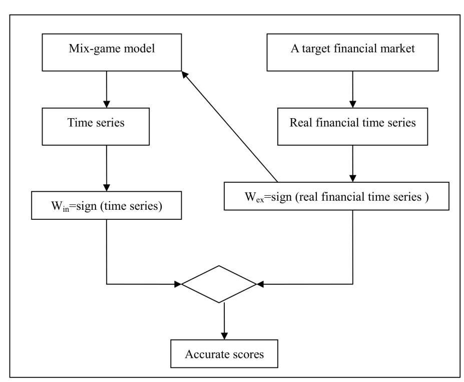
4.1 Taxonomy and Frameworks for AI Agents in Trading

AI agents in trading are governed by diverse methodologies aimed at optimizing decision-making and market dynamics. The Physiology-Driven Empathic LLM (EmLLM) integrates real-time physiological data to enhance LLM interactions, fostering empathy and improving trading strategies [60]. The noisy semi-supervised Mixture of Experts (MoE) model employs least trimmed squares estimation to manage misaligned data, enhancing robustness in dynamic markets [39]. Insights into AI robustness and neurodegeneration, termed 'neural erosion,' further inform trading frameworks [49].

The GraphAgent-Reasoner, a multi-agent framework, enhances problem-solving in trading by collaboratively addressing graph reasoning challenges [61]. This approach is crucial for processing extensive financial data. LLMs are employed to explore market equilibrium and welfare maximization, providing a foundation for understanding AI agents' roles in trading [26]. Additionally, computational models simulating imaginary and symbolic identifications advance our understanding of self-awareness in AI, crucial for complex decision-making [6].



(a) A diagram of a multi-agent reinforcement learning environment[20]



(b) A Flowchart for Predicting Financial Market Trends[22]

Figure 5: Examples of Taxonomy and Frameworks for AI Agents in Trading

Figure 5 illustrates the integration of AI and autonomous decision-making in trading. The first diagram shows a multi-agent reinforcement learning (MARL) environment where AI agents optimize actions based on state information and rewards, fostering a robust understanding of market conditions. The second diagram outlines a flowchart for predicting market trends using a mix-game model rooted in game theory, enabling AI agents to generate predictions from real financial time series [20, 22].

4.2 Multi-Agent Systems and Market Dynamics

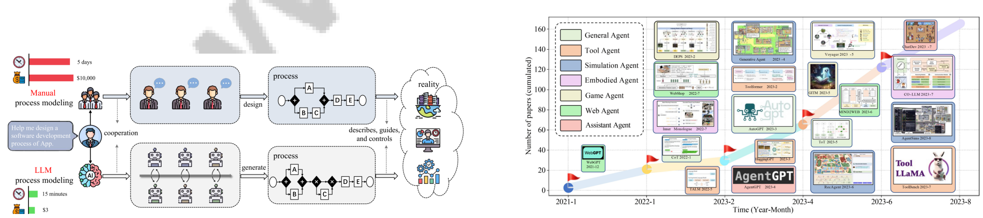
Multi-agent systems (MAS) facilitate complex interactions among autonomous agents, enhancing decision-making and optimizing market outcomes. The GraphAgent-Reasoner exemplifies MAS's impact on reasoning capabilities and decision-making in financial markets [61]. Advanced models like fine-tuned Llama 3.1 and GPT-4o enable seamless communication and coordination among agents, improving market operations' efficiency [2, 62].

In trading systems, Auto-CoT automates reasoning chain generation, boosting agents' ability to execute complex tasks and ensuring interpretability in decision-making [13, 63]. Self-organized agents (SoA) collaborate to construct comprehensive codebases, enhancing scalability in evolving market scenarios. MAS can simulate opinion dynamics within networks using LLMs, providing insights into market sentiment and behavior [64, 23, 26]. This capability is essential for predicting market trends and enabling proactive decision-making. Frameworks like Dynasaur ensure agents remain responsive to market changes through real-time adaptation and learning.

4.3 Collaborative and Adaptive Agent Frameworks

Collaborative and adaptive agent frameworks are crucial for enhancing agent-based trading systems' efficiency. Self-organized Agents (SoA) dynamically scale to manage large codebases efficiently, adapting to problem complexity [34]. Grounded Decoding (GD) allows agents to consider LLM predictions alongside grounded model probabilities, adapting actions based on real-time market conditions [65].

Role-playing frameworks enable agents to autonomously complete tasks based on initial human input, enhancing responsiveness to market trends [66]. Integrating Lacanian psychoanalysis with active inference offers a novel framework for self-identification in intelligent agents, allowing them to simulate human-like decision-making processes [6].



(a) Efficient Software Development Process Design Using Artificial Intelligence and Human Collaboration[67]

(b) Timeline of Agent Research and Development[37]

Figure 6: Examples of Collaborative and Adaptive Agent Frameworks

Figure 7 illustrates the key components of collaborative and adaptive agent frameworks, highlighting self-organized agents for dynamic scaling, grounded decoding for real-time adaptation, and role-playing frameworks for autonomous task completion and human-like decision-making. The first example visualizes a framework where AI and human expertise converge to enhance software development processes. The second example categorizes advancements in agent technologies from 2021 to 2023, showcasing the dynamic and adaptive nature of agent frameworks across various domains [67, 37].

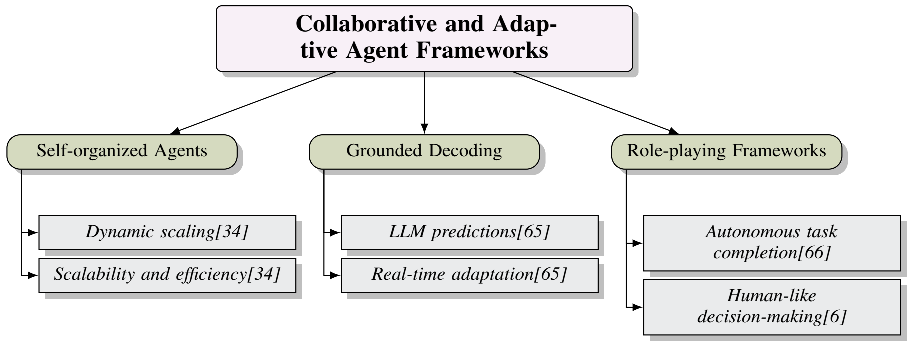


Figure 7: This figure illustrates the key components of collaborative and adaptive agent frameworks, highlighting self-organized agents for dynamic scaling, grounded decoding for real-time adaptation, and role-playing frameworks for autonomous task completion and human-like decision-making.

5 Market Microstructure and Algorithmic Trading

5.1 Theoretical Foundations of Market Microstructure

Understanding market microstructure is crucial for elucidating trading processes, price formation, and liquidity in financial markets. This area examines the interactions of diverse agents, which significantly influence market dynamics and efficiency. The SYMBA model, for instance, effectively captures unique features of crypto markets, such as non-normal returns and volume clustering, highlighting the distinct characteristics of decentralized markets [20]. By integrating agent-based models with traditional frameworks like the GARCH model, a comprehensive understanding of financial market complexities emerges, accounting for diverse participant interactions that contribute to volatility and unpredictability [21]. These models' modularity allows for the incorporation of historical knowledge, enhancing market predictions and informing trading strategies [68].

Large language models (LLMs) play a pivotal role in this context, with theoretical explorations emphasizing trustworthiness and quality analysis, both critical to trading decisions and market behavior [69]. LLMs' ability to dynamically generate rules and adapt to evolving temporal knowledge graphs ensures trading rules' effectiveness amid changing conditions [70]. Market impact studies further categorize effects into individual transaction impacts, aggregate transaction effects, and hidden orders' influence on price dynamics, offering insights for optimizing trading strategies and risk management [19].

Despite these advancements, current LLMs face limitations in replicating adaptive and psychological factors affecting human trading behavior, underscoring the need for further research. These limitations can hinder trade execution effectiveness, highlighting the importance of incorporating psychological insights into model design [26]. As the field evolves, integrating advanced AI technologies like LLMs with traditional market microstructure theories promises to enhance financial markets' understanding and functionality, fostering innovation and efficiency in trading systems [11].

5.2 Technological Innovations in Trading Systems

Technological innovations have significantly enhanced the efficiency and effectiveness of financial markets through advanced algorithms and AI. The integration of Large Language Models (LLMs) into trading systems exemplifies this progress, facilitating sophisticated data analysis and decision-making. The FLAME framework, for example, enhances taxonomy expansion tasks by combining few-shot prompting with reinforcement learning, effectively leveraging LLM capabilities to deepen market dynamics understanding and improve trading strategies' adaptability [52].

Methods like MoDeGPT highlight modular optimization's potential, achieving compression rates of 25-30

Additionally, the Event Detection Framework (EDF) integrates LLMs with clustering algorithms to identify and categorize news events from the GDELT database, providing a robust mechanism for isolating market-relevant information and guiding trading decisions [27]. This integration illustrates

LLMs' role in augmenting traditional data analysis techniques, enhancing market insights' accuracy and timeliness. The asynchronous tool usage method, which processes multiple events concurrently, further showcases innovations by enabling fluid interactions and timely user query responses [62], a capability vital for maintaining agility in fast-paced market environments.

Moreover, the PriDe framework represents a significant advancement in LLM performance by addressing selection bias without requiring sample labels [29], thereby enhancing trading model predictions' accuracy and improving trading outcomes' reliability.

6 Natural Language Processing in Financial Forecasting

6.1 Techniques for Processing Financial Textual Data

Extracting actionable insights from financial textual data is pivotal for strategic decision-making in markets. Advances in Large Language Models (LLMs) have enhanced natural language understanding and generation, achieving high accuracy across various applications. These models excel in processing complex language constructs, crucial for summarizing information from intricate sources such as financial reports and rapidly evolving news articles, where timely event detection is essential [8, 27]. LLMs support keyword extraction, text embedding, and event summarization, thus improving news event clustering and financial textual data analysis [27]. Tarekegn et al. highlight the potential of LLMs to analyze extensive financial datasets, facilitating more accurate and timely decision-making.

Innovative methods, like the graph-based approach utilized by the GraphAgent-Reasoner, enhance financial textual data analysis by leveraging structural relationships, offering deeper insights into interconnected financial information [61]. The Dynamic Adaptive Optimization (DAO) framework exemplifies LLM application in financial contexts, evaluating performance on a dataset of 23,242 preprocessed news and analysis texts related to EUR/USD exchange rates, underscoring the necessity of comprehensive datasets for training LLMs to accurately capture financial language nuances [40].

6.2 Advancements in NLP Techniques for Financial Insights

Advancements in Natural Language Processing (NLP) have significantly enhanced the extraction of financial insights, driven by innovations in LLMs and related technologies. The integration of multimodal data, including audio and text, allows for a comprehensive analysis of financial information, leading to deeper insights into market dynamics [53]. Models like InternVL2-8B-MPO demonstrate enhanced multimodal reasoning and reduced hallucinations, showcasing the effectiveness of preference optimization techniques [25].

LLMs' ability to interpret complex human language, including metaphors and nuanced expressions, has advanced the extraction of detailed insights from financial texts. Incorporating physiological data into LLMs represents a significant advancement, enabling the extraction of insights related to users' psychological states, crucial for understanding market sentiments [60]. The emergence of Retrieval-Augmented Generation (RAG) technology enhances LLM performance in real-world applications by improving the retrieval and utilization of relevant information [71]. Integrating LLMs with Knowledge Graphs (KGs) is a promising direction for enhancing interpretability and facilitating financial insight extraction [59].

Advancements in structured pruning techniques, such as LoRAShear, demonstrate effectiveness in maintaining model performance while reducing computational costs, achieving only a 1% performance degradation at a 20% pruning ratio [24]. These techniques are vital for the efficient deployment of LLMs in financial applications, where computational resources are often constrained [18]. The CMAT framework illustrates that small-parameter models can achieve performance comparable to larger models when optimized effectively, highlighting the potential for more efficient NLP techniques [58]. Additionally, advancements in tool learning with LLMs emphasize the importance of systematic implementation and comprehensive evaluation methods to assess tool usage and effectiveness [72].

7 Integration and Challenges

The integration of Large Language Models (LLMs) into financial markets offers substantial opportunities alongside significant challenges. Critical analysis of interdisciplinary collaboration and ethical frameworks is vital for addressing the complex issues posed by LLMs. Understanding these aspects enables leveraging LLMs' potential while addressing ethical concerns during deployment.

7.1 Interdisciplinary Collaboration and Ethical Frameworks

Effective integration of LLMs in financial markets requires robust interdisciplinary collaboration and ethical frameworks. The CREATOR framework emphasizes tool creation to foster AI development collaboration and ensure LLMs' effective integration across applications [73]. Such collaboration is crucial for overcoming limitations like catastrophic forgetting and adapting to rapidly changing information [28]. Interdisciplinary efforts enhance systems like MetaGPT, designed for dynamic environments and complex software scenarios [36]. Ethical frameworks guide responsible LLM deployment to mitigate societal risks from emergent behaviors, particularly in financial markets where biased AI decisions can have significant impacts [74].

Gkatzia et al.'s benchmark provides structured evaluation methods for data presentation techniques in financial contexts, offering insights for decision support systems [17]. Future research should focus on expanding datasets and transitioning to GPU systems to enhance computational capabilities, facilitating comparisons with classical models and improving LLM performance [41]. As LLM capabilities evolve, efficient training methods and exploration of ethical implications are necessary. Investigating emergent abilities' factors is crucial for safe and effective model utilization [46]. Bridging psychology and AI through interdisciplinary approaches can deepen understanding of self-identification in machines, revealing LLMs' potential and limitations [6].

7.2 Computational Resource Challenges

Deploying LLMs in financial markets presents significant computational resource challenges. These models' complexity and scale demand substantial computational power for training and inference. High resource requirements and limited training datasets hinder existing pruning methods' effectiveness [24]. Increased computational costs may adversely affect inference times, posing challenges for real-time financial applications [27]. Efficient inference processes are essential for addressing these challenges. Models like GraphAgent-Reasoner, handling graphs with up to 1,000 nodes, address computational issues in deploying complex reasoning models [61]. However, studies often inadequately address computational costs and model generalization limitations across modalities [4].

The Dynamic Adaptive Optimization (DAO) method illustrates potential computational overhead during training, indicating resource management challenges [40]. The Noisy Mixture of Experts (Noisy MoE) method may underperform under high noise levels, emphasizing the need for robust computational strategies [39]. Future research should improve Knowledge Graph-enhanced LLMs (KGLLMs), explore multimodal and temporal Knowledge Graphs (KGs), and develop methods for better knowledge integration and interpretability [42]. These efforts are crucial for effective and sustainable AI operations in the fast-paced, resource-intensive financial markets.

7.3 Data Privacy and Security

Data privacy and security are paramount in deploying LLMs within financial applications due to sensitive financial data and potential misuse. Carlini et al.'s framework evaluates privacy risks associated with LLMs, aiding in developing robust privacy-preserving techniques [75]. This framework is vital for identifying vulnerabilities and preventing LLMs from inadvertently exposing sensitive information. Integrating LLMs with financial systems raises concerns about synthetic data generation and handling. While synthetic data enhances model training, it risks producing biased datasets, challenging diversity and representativeness [45]. Addressing these challenges is crucial for maintaining data integrity and trust.

Privacy protection capabilities must align with LLM training guidelines to mitigate sensitive data risks [76]. Adhering to privacy standards is vital for safeguarding user information and ensuring regulatory compliance. Jin et al.'s survey highlights gaps in addressing democratization and privacy challenges, underscoring the need for comprehensive strategies to protect data privacy in financial markets [3]. The Viz system presents a legally compliant approach to data privacy by pre-training LLMs on non-copyrighted datasets and offering fine-tuned models in a legal marketplace [32]. This approach ensures LLMs are trained and deployed respecting intellectual property rights and data privacy regulations.

High resource requirements for model training complicate the privacy landscape, necessitating vast data collection and processing [11]. Ensuring data privacy and security in such environments requires innovative solutions. Future work should address privacy concerns related to sensitive physiological data collection, particularly in financial applications where such data may enhance interactions and decision-making [60]. Transparency and user awareness regarding AI interactions' privacy are emphasized, highlighting the importance of clear communication and informed consent [30].

7.4 Model Interpretability and Bias

Interpretability and bias are significant challenges in deploying LLMs within financial markets, where opaque decision-making and biased outputs can have substantial consequences. The "black-box" nature of LLMs complicates accountability, making it difficult for stakeholders to understand decision-making processes [50]. This lack of transparency can lead to discrimination and misinformation, as biases in training data are propagated. Efforts to enhance model interpretability, such as the LUNA framework, aim to improve understanding of LLM behaviors by providing insights into decision-making processes [69]. However, challenges remain in capturing all relevant information, particularly in complex scenarios like earnings conference calls [10].

Bias is compounded by issues such as position bias and numeric hallucinations, impacting financial analyses' accuracy and reliability [8]. The ECC Analyzer's limitations in extracting comprehensive insights from complex datasets underscore the struggle for unbiased outputs in niche domains [10]. Machine unlearning mechanisms offer a promising approach to enhancing LLMs' ethical and legal robustness by enabling biased or incorrect information removal [77]. This approach addresses interpretability and bias challenges, ensuring LLMs align with ethical standards and legal requirements.

The critical aspect of utility assessment in decision-making underscores model interpretability's importance, as it relates to AI systems' ability to make informed, unbiased decisions [16]. Effective benchmarks for evaluating privacy risks and model biases are emphasized, highlighting the need to understand and mitigate these challenges for responsible LLM deployment in financial applications [75].

8 Conclusion

8.1 Future Directions and Research Opportunities

The horizon for Large Language Models (LLMs) in financial markets is expansive, presenting numerous opportunities to enhance their capabilities and address existing challenges. One promising avenue is the development of lightweight models optimized for edge deployment, which can improve performance in complex network environments and address multimodal alignment challenges. A critical area of focus is enhancing LLMs' capacity to conduct independent research and assimilate new information effectively. By refining these models with real-world discourse data, their ability to simulate diverse human beliefs and behaviors can be broadened, enhancing their effectiveness in modeling opinion dynamics within networks. Additionally, improving the transferability of priors across domains and strengthening model robustness against selection bias are essential for ensuring LLM reliability across various applications.

In the realm of workflow automation, exploring alternative learning paradigms, such as teacher-student or adversarial learning, could enhance collaborative learning between workflow generators and interpreter LLMs. Expanding the CREATOR framework across different domains and investigating LLMs as tool creators present significant opportunities for exploration. Improving model scalability through multi-GPU configurations and customizing implementations for specific datasets can further enhance LLM efficiency in financial contexts. Furthermore, innovative strategies to mitigate forgetting, improve temporal generalization, and develop adaptive architectures are crucial for maintaining model performance over time.

Future research should delve into the interplay between uncertainty, risk-taking behavior, and gender differences in decision-making, offering valuable insights for tailoring LLM applications in financial markets. Applying CMAT in broader contexts and diverse task types can enhance the versatility of multi-agent collaboration frameworks. Investigating the mechanisms behind emergent abilities, developing new architectures, and improving data quality to lower the scale threshold for emergent capabilities are critical areas for further inquiry. Integrating LLMs with external tools and retrieval systems, while refining training techniques, will be pivotal in aligning model outputs with human preferences. Moreover, enhancing the reliability of reasoning paths and inducing similar reasoning capabilities in smaller models remain essential.

Strengthening the robustness of watermarks against adversarial attacks and improving detection methods for low-entropy text are vital research areas. Further refinements to the MPO method and strategies for enhancing multimodal reasoning could significantly advance the field. Investigating alternative noise distributions and validating findings against real-world cognitive decline data could bolster the robustness of the neural erosion concept. Future research should focus on optimizing the pruning process and applying LoRAShear to a wider range of LLM architectures. Refining privacy metrics and exploring additional privacy-preserving techniques for LLM training processes are crucial for protecting sensitive information. Addressing current model limitations and emerging trends in multimodal AI will further enhance LLM capabilities. Finally, refining GAR algorithms and applying them to complex real-world graph reasoning problems will open new avenues in AI and financial forecasting.

These research directions aim to elevate the analytical capabilities of LLMs, ensuring they provide valuable insights and support strategic decision-making in the evolving financial landscape. Future efforts will prioritize improving the quality of extracted factors and the reproducibility of results in financial time-series forecasting, while exploring diverse market conditions and enhanced training protocols for LLMs. Advanced AI techniques for monitoring event progression and evaluating the Cluster Stability Assessment Index (CSAI) across various domains will further drive innovation. Lastly, refining models to incorporate nuanced psychological factors and exploring practical applications in mental health diagnostics could unveil new research and application avenues.

8.2 Challenges and Knowledge Gaps

The implementation of Large Language Models (LLMs) in financial markets presents several challenges and knowledge gaps that require further investigation. A primary challenge is the systematic biases in language model responses, which can lead to flawed conclusions regarding their alignment with human populations. These biases significantly impact decision-making processes and the reliability of insights derived from LLMs, highlighting the need for sophisticated bias detection and mitigation strategies.

Another critical challenge involves designing and implementing new business models enabled by LLMs. Future research should quantify the economic value generated by these transformation archetypes, providing clearer insights into the potential benefits and limitations of integrating LLMs into various financial applications. This exploration is essential for maximizing the economic impact of LLMs and ensuring their deployment aligns with strategic business objectives.

Model interpretability remains a significant knowledge gap, as the "black-box" nature of LLMs complicates understanding their decision-making processes. Enhancing transparency and accountability in LLM outputs is vital for building stakeholder trust and effective integration into financial systems. Additionally, the scalability and computational resource requirements of LLMs pose ongoing challenges, necessitating the development of efficient training and deployment strategies for practicality in resource-constrained environments.

Furthermore, integrating multimodal data sources with LLMs presents opportunities for comprehensive analyses but introduces complexities related to data alignment and processing. Addressing these challenges requires advancements in multimodal learning techniques and robust frameworks for seamlessly integrating diverse data types.

References

[1] Bosheng Ding, Chengwei Qin, Ruochen Zhao, Tianze Luo, Xinze Li, Guizhen Chen, Wenhan Xia, Junjie Hu, Anh Tuan Luu, and Shafiq Joty. Data augmentation using large language models: Data perspectives, learning paradigms and challenges, 2024.

[2] Hao Kang and Chenyan Xiong. Researcharena: Benchmarking large language models' ability to collect and organize information as research agents, 2025.

[3] Yizhang Jin, Jian Li, Yexin Liu, Tianjun Gu, Kai Wu, Zhengkai Jiang, Muyang He, Bo Zhao, Xin Tan, Zhenye Gan, Yabiao Wang, Chengjie Wang, and Lizhuang Ma. Efficient multimodal large language models: A survey, 2024.

[4] Soyeon Caren Han, Feiqi Cao, Josiah Poon, and Roberto Navigli. Multimodal large language models and tunings: Vision, language, sensors, audio, and beyond, 2024.

[5] Zelong Li, Shuyuan Xu, Kai Mei, Wenyue Hua, Balaji Rama, Om Raheja, Hao Wang, He Zhu, and Yongfeng Zhang. Autoflow: Automated workflow generation for large language model agents, 2024.

[6] Lingyu Li and Chunbo Li. Enabling self-identification in intelligent agent: insights from computational psychoanalysis, 2024.

[7] Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564*, 2023.

[8] Tianyu Cao, Natraj Raman, Danial Dervovic, and Chenhao Tan. Characterizing multimodal long-form summarization: A case study on financial reports, 2024.

[9] Meiyun Wang, Kiyoshi Izumi, and Hiroki Sakaji. Llmfactor: Extracting profitable factors through prompts for explainable stock movement prediction, 2024.

[10] Yupeng Cao, Zhi Chen, Qingyun Pei, Nathan Jinseok Lee, K. P. Subbalakshmi, and Papa Momar Ndiaye. Ecc analyzer: Extract trading signal from earnings conference calls using large language model for stock performance prediction, 2024.

[11] Sifan Long, Fengxiao Tang, Yangfan Li, Tiao Tan, Zhengjie Jin, Ming Zhao, and Nei Kato. 6g comprehensive intelligence: network operations and optimization based on large language models, 2025.

[12] Hao Zhou, Chengming Hu, Ye Yuan, Yufei Cui, Yili Jin, Can Chen, Haolun Wu, Dun Yuan, Li Jiang, Di Wu, Xue Liu, Charlie Zhang, Xianbin Wang, and Jiangchuan Liu. Large language model (llm) for telecommunications: A comprehensive survey on principles, key techniques, and opportunities, 2024.

[13] Zhuosheng Zhang, Yao Yao, Aston Zhang, Xiangru Tang, Xinbei Ma, Zhiwei He, Yiming Wang, Mark Gerstein, Rui Wang, Gongshen Liu, et al. Igniting language intelligence: The hitchhiker's guide from chain-of-thought reasoning to language agents. *arXiv preprint arXiv:2311.11797*, 2023.

[14] Scott Reed, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexander Novikov, Gabriel Barth-Maron, Mai Gimenez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, et al. A generalist agent. *arXiv preprint arXiv:2205.06175*, 2022.

[15] Hamed Babaei Giglou, Jennifer D'Souza, and Sören Auer. Llms4ol: Large language models for ontology learning, 2023.

[16] Gianluca Bontempi. Between accurate prediction and poor decision making: the ai/ml gap, 2023.

[17] Dimitra Gkatzia, Oliver Lemon, and Verena Rieser. Natural language generation enhances human decision-making with uncertain information, 2016.

[18] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. A comprehensive overview of large language models. *arXiv preprint arXiv:2307.06435*, 2023.

[19] Jean-Philippe Bouchaud, J. Doyne Farmer, and Fabrizio Lillo. How markets slowly digest changes in supply and demand, 2008.

[20] Johann Lussange, Stefano Vrizzi, Stefano Palminteri, and Boris Gutkin. Modelling crypto markets by multi-agent reinforcement learning, 2024.

[21] Kei Nakagawa, Masanori Hirano, Kentaro Minami, and Takanobu Mizuta. A multi-agent market model can explain the impact of ai traders in financial markets – a new microfoundations of garch model, 2024.

[22] Chengling Gou. Deduction of initial strategy distributions of agents in mix-game model, 2006.

[23] Yajing Wang and Zongwei Luo. Enhance multi-domain sentiment analysis of review texts through prompting strategies, 2024.

[24] Tianyi Chen, Tianyu Ding, Badal Yadav, Ilya Zharkov, and Luming Liang. Lorashear: Efficient large language model structured pruning and knowledge recovery, 2023.

[25] Weiyun Wang, Zhe Chen, Wenhai Wang, Yue Cao, Yangzhou Liu, Zhangwei Gao, Jinguo Zhu, Xizhou Zhu, Lewei Lu, Yu Qiao, and Jifeng Dai. Enhancing the reasoning ability of multimodal large language models via mixed preference optimization, 2024.

[26] Jingru Jia and Zehua Yuan. An experimental study of competitive market behavior through llms, 2024.

[27] Adane Nega Tarekegn. Large language model enhanced clustering for news event detection, 2024.

[28] Junhao Zheng, Shengjie Qiu, Chengming Shi, and Qianli Ma. Towards lifelong learning of large language models: A survey, 2024.

[29] Chujie Zheng, Hao Zhou, Fandong Meng, Jie Zhou, and Minlie Huang. Large language models are not robust multiple choice selectors. In *The Twelfth International Conference on Learning Representations*, 2023.

[30] Imdad Ullah, Najm Hassan, Sukhpal Singh Gill, Basem Suleiman, Tariq Ahamed Ahanger, Zawar Shah, Junaid Qadir, and Salil S. Kanhere. Privacy preserving large language models: Chatgpt case study based vision and framework, 2023.

[31] Sen Huang, Kaixiang Yang, Sheng Qi, and Rui Wang. When large language model meets optimization, 2024.

[32] Dipankar Sarkar. Viz: A qlora-based copyright marketplace for legally compliant generative ai, 2023.

[33] Ayush Chopra, Shashank Kumar, Nurullah Giray-Kuru, Ramesh Raskar, and Arnau Quera-Bofarull. On the limits of agency in agent-based models, 2024.

[34] Yoichi Ishibashi and Yoshimasa Nishimura. Self-organized agents: A llm multi-agent framework toward ultra large-scale code generation and optimization, 2024.

[35] Dang Nguyen, Viet Dac Lai, Seunghyun Yoon, Ryan A. Rossi, Handong Zhao, Ruiyi Zhang, Puneet Mathur, Nedim Lipka, Yu Wang, Trung Bui, Franck Dernoncourt, and Tianyi Zhou. Dynasaur: Large language agents beyond predefined actions, 2024.

[36] Sirui Hong, Xiawu Zheng, Jonathan Chen, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, et al. Metagpt: Meta programming for multi-agent collaborative framework. *arXiv preprint arXiv:2308.00352*, 2023.

[37] Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, et al. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6):186345, 2024.

[38] Paul Bilokon, Oleksandr Bilokon, and Saeed Amen. A compendium of data sources for data science, machine learning, and artificial intelligence, 2023.

[39] Oh-Ran Kwon, Gourab Mukherjee, and Jacob Bien. Semi-supervised learning of noisy mixture of experts models, 2024.

[40] Hongcheng Ding, Xuanze Zhao, Shamsul Nahar Abdullah, Deshinta Arrova Dewi, Zixiao Jiang, and Xiangyu Shi. Dynamic adaptive optimization for effective sentiment analysis fine-tuning on large language models, 2024.

[41] Hala Hawashin and Mehrnoosh Sadrzadeh. Multimodal structure-aware quantum data processing, 2025.

[42] Linyao Yang, Hongyang Chen, Zhao Li, Xiao Ding, and Xindong Wu. Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling, 2024.

[43] Xinyi Mou, Jingcong Liang, Jiayu Lin, Xinnong Zhang, Xiawei Liu, Shiyue Yang, Rong Ye, Lei Chen, Haoyu Kuang, Xuanjing Huang, and Zhongyu Wei. Agentsense: Benchmarking social intelligence of language agents through interactive scenarios, 2024.

[44] Xinyi Hou, Yanjie Zhao, Yue Liu, Zhou Yang, Kailong Wang, Li Li, Xiapu Luo, David Lo, John Grundy, and Haoyu Wang. Large language models for software engineering: A systematic literature review. *ACM Transactions on Software Engineering and Methodology*, 33(8):1–79, 2024.

[45] Xu Guo and Yiqiang Chen. Generative ai for synthetic data generation: Methods, challenges and the future, 2024.

[46] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.

[47] Jie Huang and Kevin Chen-Chuan Chang. Towards reasoning in large language models: A survey. *arXiv preprint arXiv:2212.10403*, 2022.

[48] Q. Vera Liao and Jennifer Wortman Vaughan. Ai transparency in the age of llms: A human-centered research roadmap, 2023.

[49] Antonios Alexos, Yu-Dai Tsai, Ian Domingo, Maryam Pishgar, and Pierre Baldi. Neural erosion: Emulating controlled neurodegeneration and aging in ai systems, 2024.

[50] Junfeng Jiao, Saleh Afroogh, Yiming Xu, and Connor Phillips. Navigating llm ethics: Advancements, challenges, and future directions, 2024.

[51] Chi-Heng Lin, Shangqian Gao, James Seale Smith, Abhishek Patel, Shikhar Tuli, Yilin Shen, Hongxia Jin, and Yen-Chang Hsu. Modegpt: Modular decomposition for large language model compression, 2025.

[52] Sahil Mishra, Ujjwal Sudev, and Tanmoy Chakraborty. Flame: Self-supervised low-resource taxonomy expansion using large language models, 2024.

[53] Zichen Zhu, Yang Xu, Lu Chen, Jingkai Yang, Yichuan Ma, Yiming Sun, Hailin Wen, Jiaqi Liu, Jinyu Cai, Yingzi Ma, Situo Zhang, Zihan Zhao, Liangtai Sun, and Kai Yu. Multi: Multimodal understanding leaderboard with text and images, 2025.

[54] Chunyi Sun, Junlin Han, Weijian Deng, Xinlong Wang, Zishan Qin, and Stephen Gould. 3d-gpt: Procedural 3d modeling with large language models, 2024.

[55] Baha Rababah, Shang, Wu, Matthew Kwiatkowski, Carson Leung, and Cuneyt Gurcan Akcora. Sok: Prompt hacking of large language models, 2024.

[56] John Kirchenbauer, Jonas Geiping, Yuxin Wen, Jonathan Katz, Ian Miers, and Tom Goldstein. A watermark for large language models. In *International Conference on Machine Learning*, pages 17061–17084. PMLR, 2023.

[57] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

[58] Xuechen Liang, Meiling Tao, Yinghui Xia, Tianyu Shi, Jun Wang, and JingSong Yang. Cmat: A multi-agent collaboration tuning framework for enhancing small language models, 2024.

[59] Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[60] Poorvesh Dongre, Majid Behravan, Kunal Gupta, Mark Billinghurst, and Denis Gračanin. Integrating physiological data with large language models for empathic human-ai interaction, 2024.

[61] Yuwei Hu, Runlin Lei, Xinyi Huang, Zhewei Wei, and Yongchao Liu. Scalable and accurate graph reasoning with llm-based multi-agents, 2024.

[62] Antonio A. Ginart, Naveen Kodali, Jason Lee, Caiming Xiong, Silvio Savarese, and John Emmons. Asynchronous tool usage for real-time agents, 2024.

[63] Zhuosheng Zhang, Aston Zhang, Mu Li, and Alex Smola. Automatic chain of thought prompting in large language models. *arXiv preprint arXiv:2210.03493*, 2022.

[64] Yun-Shiuan Chuang, Agam Goyal, Nikunj Harlalka, Siddharth Suresh, Robert Hawkins, Sijia Yang, Dhavan Shah, Junjie Hu, and Timothy T. Rogers. Simulating opinion dynamics with networks of llm-based agents, 2024.

[65] Wenlong Huang, Fei Xia, Dhruv Shah, Danny Driess, Andy Zeng, Yao Lu, Pete Florence, Igor Mordatch, Sergey Levine, Karol Hausman, and Brian Ichter. Grounded decoding: Guiding text generation with grounded models for embodied agents, 2023.

[66] Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. Camel: Communicative agents for "mind" exploration of large language model society, 2023.

[67] Leilei Lin, Yumeng Jin, Yingming Zhou, Wenlong Chen, and Chen Qian. Mao: A framework for process model generation with multi-agent orchestration, 2024.

[68] Chidaksh Ravuru, Sagar Srinivas Sakhinana, and Venkataramana Runkana. Agentic retrieval-augmented generation for time series analysis, 2024.

[69] Da Song, Xuan Xie, Jiayang Song, Derui Zhu, Yuheng Huang, Felix Juefei-Xu, and Lei Ma. Luna: A model-based universal analysis framework for large language models, 2024.

[70] Jiapu Wang, Kai Sun, Linhao Luo, Wei Wei, Yongli Hu, Alan Wee-Chung Liew, Shirui Pan, and Baocai Yin. Large language models-guided dynamic adaptation for temporal knowledge graph reasoning, 2024.

[71] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2023.

[72] Changle Qu, Sunhao Dai, Xiaochi Wei, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, Jun Xu, and Ji-Rong Wen. Tool learning with large language models: A survey, 2024.

[73] Cheng Qian, Chi Han, Yi R. Fung, Yujia Qin, Zhiyuan Liu, and Heng Ji. Creator: Tool creation for disentangling abstract and concrete reasoning of large language models, 2024.

[74] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*, 2022.

[75] Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, et al. Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 2633–2650, 2021.

[76] Yuqi Yang, Xiaowen Huang, and Jitao Sang. Exploring the privacy protection capabilities of chinese large language models, 2024.

[77] Saaketh Koundinya Gundavarapu, Shreya Agarwal, Arushi Arora, and Chandana Thimmalapura Jagadeeshaiah. Machine unlearning in large language models, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.