SCIENCE CHINA
Information Sciences

• REVIEW •

July 2026, Vol. 69, Iss. 7, 171201:1–171201:22
https://doi.org/10.1007/s11432-026-4986-x

Large language model-based multi-agent systems for financial markets simulation: a survey

Qinyuan LIU[1,2], Lihang YAO[1,2], Zidong WANG[3,4*], Yufan YANG[1,2], Yifei TANG[1,2], Dawei CHENG[1,2] & Changjun JIANG[1,2]

[1] *Department of Computer Science and Technology, Tongji University, Shanghai 201804, China*
[2] *Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 200092, China*
[3] *College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China*
[4] *Department of Computer Science, Brunel University of London, Uxbridge UB8 3PH, UK*

Received 5 January 2026/Revised 4 March 2026/Accepted 9 April 2026/Published online 24 June 2026

Abstract Traditional financial modeling approaches, such as econometric methods and dynamic stochastic general equilibrium models, exhibit inherent limitations in capturing the highly interactive and strategic nature of modern financial systems. To overcome these limitations, multi-agent systems (MASs) have increasingly been adopted as a computational paradigm well suited for simulating the complex behaviors and interactions of participants in financial environments. Within this paradigm, the emerging class of large language model (LLM)-based MASs has demonstrated unprecedented potential for modeling intricate financial interactions. This study provides a systematic review of LLM-based MAS applications in financial markets. We first present the motivation for adopting LLM-based MASs and highlight their key characteristics, including heterogeneity, autonomy, adaptability, and bounded rationality, which render them particularly effective for representing complex financial ecosystems. Then, we conduct a technical analysis of financial agents, providing the enabling techniques for LLM-based agents, a taxonomy of their foundational models, and the underlying mechanisms that facilitate the emergence of collective intelligence in LLM-based MASs. Furthermore, we survey some representative applications of LLM-based MASs in financial markets, including market dynamics simulation, systemic risk analysis, policy evaluation, algorithmic trading, and sentiment analysis, while critically discussing the associated technical and regulatory challenges. Through this comprehensive synthesis, we aim to provide a state-of-the-art reference for intelligent agent-based approaches in computational finance and to identify key opportunities and limitations that characterize this rapidly evolving research field.

Keywords large language model, multi-agent system, financial markets, collective intelligence, market dynamics simulation, systemic risk analysis, policy evaluation, sentiment analysis

Citation Liu Q Y, Yao L H, Wang Z D, et al. Large language model-based multi-agent systems for financial markets simulation: a survey. Sci China Inf Sci, 2026, 69(7): 171201, https://doi.org/10.1007/s11432-026-4986-x

1 Introduction

Financial markets operate as complex systems. Within these systems, systemic complexity, pervasive uncertainty, and ongoing structural transformation emerge from the strategic interactions of heterogeneous agents embedded in interconnected networks. These features give rise to nonlinear dynamical behaviors that frequently challenge the traditional financial modeling approaches, such as the econometric time-series models and the dynamic stochastic general equilibrium (DSGE) frameworks [1]. In response, advances in computational methodologies have fostered the interdisciplinary incorporation of novel modeling paradigms into financial theory. This integration enables more effective analysis across a wide range of financial tasks [2].

Recent advances in computing power and data availability have enabled large-scale agent-based simulations, thereby increasing the academic interest in applying multi-agent systems (MASs) to the financial domain [3]. MAS adopts a bottom-up modeling paradigm characterized by heterogeneous agents and rich interaction structures. This approach allows for the emergence of nonlinear macro-level dynamics, which are difficult to capture using traditional equilibrium-based models. This modeling approach is referred to by different terminologies across disciplines, such as agent-based modeling (ABM) in finance and individual-based modeling (IBM) in ecology. Specifically, an LLM-based agent is a special type of intelligent agent driven by a large language model (LLM).

A notable recent development is the integration of LLM with MAS. Unlike general text, financial data feature low signal-to-noise ratios, strict temporal dependencies, and a tight coupling of quantitative metrics and qualitative

* Corresponding author (email: Zidong.Wang@brunel.ac.uk)

© Science China Press 2026

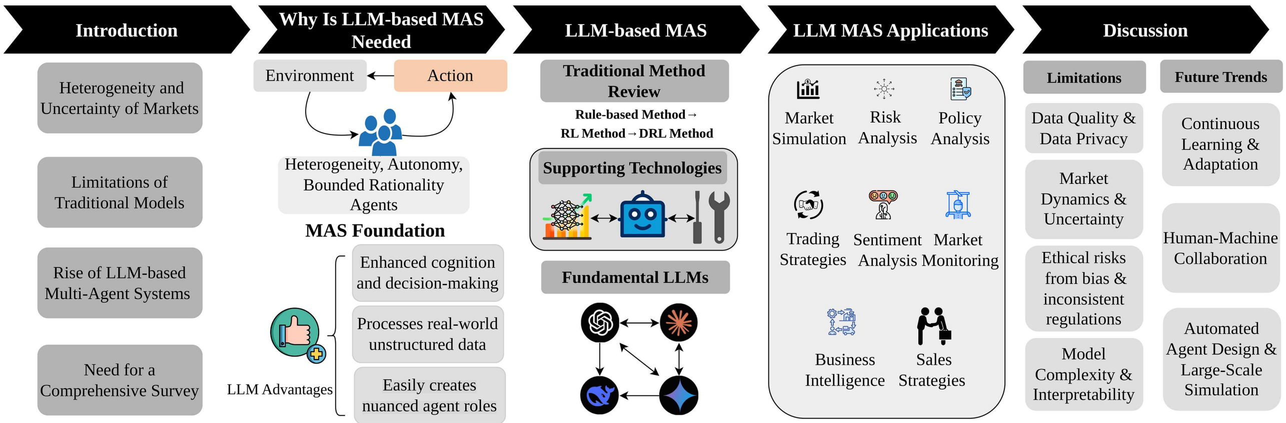

Figure 1 (Color online) A roadmap of this survey, outlining its logical structure. The survey begins by motivating the adoption of LLM-based MAS, highlighting the limitations of traditional models and the key advantages of LLM-based MAS (left). It then delves into the technical architecture of these systems, covering foundational LLMs and supporting technologies (middle). Subsequently, the study surveys a wide range of financial applications before concluding with a discussion of the major limitations and future trends that are shaping the field (right).

Table 1 Comparison of core features between this survey and related surveys.

Ref.	General financial NLP	LLM-based agent focus	MAS focus	MAS architecture	Market simulation	Agent architecture
Li et al. [6]	✓	✓	✗	✗	✗	✓
Zhao et al. [4]	✓	✗	✗	✗	✓	✗
Nie et al. [2]	✓	✓	✗	✓	✓	✓
Ours	✓	✓	✓	✓	✓	✓

narratives. Thus, applying LLMs in finance demands specialized mechanisms to handle real-time volatility, mitigate hallucinations, and ensure rigorous reasoning to prevent severe economic consequences. In financial contexts, LLM-based MASs have been applied across several key areas. For example, they enable the analysis of financial crises and the simulation of policy interventions with enhanced interpretability. This capability helps policymakers assess the impacts of different scenarios, providing empirical foundations for informed decision-making. Moreover, LLM-based MASs have also been utilized for algorithmic trading, sentiment analysis, market monitoring, and related financial tasks [4].

A key advantage of LLM-based MASs lies in their explanatory transparency. By providing explicit narratives that link micro-level agent behaviors to macro-level system outcomes, LLM-based MASs offer a clear advantage over the traditional econometric models and machine-learning-based approaches. Existing surveys primarily focus on financial aspects while largely overlooking the underlying agent technologies [3], and the literature addressing the application of LLM-based agents in financial markets remains fragmented and incomplete [5]. Table 1 provides a quantitative comparison between this survey and existing related studies [2, 4, 6].

Therefore, this survey seeks to bridge these gaps by systematically reviewing the theoretical foundations of agents and providing a comprehensive analysis of the associated technical methodologies, application scenarios, and future challenges and prospects. To ensure a focused and high-quality review, the literature selection criteria primarily emphasize recent studies published from 2022 onwards. Specifically, we prioritize studies that explicitly integrate LLMs with multi-agent systems for financial market and economic simulation. These are selected over traditional rule-based or purely DRL-driven agent-based models. The overall structure of the article is shown in Figure 1. Our work offers a holistic perspective on LLM-based MASs for financial modeling, aiming to furnish both researchers and practitioners with a coherent and up-to-date reference for this rapidly evolving field.

2 Why is LLM-based MAS needed?

As discussed above, traditional financial models exhibit several fundamental limitations. The primary reason is that they often assume agent homogeneity, ignoring the diverse behaviors and interactions of real market participants, which restricts their ability to simulate realistic scenarios and systemic risks [7, 8]. Furthermore, these models generally presume a static environment, limiting their adaptability to the non-stationary and volatile nature of actual financial markets [9]. Finally, the paradox of rational expectations posits that agents act with perfect rationality to maximize outcomes. However, this conflicts with empirical evidence from behavioral economics. Real-world decisions are inherently made under cognitive constraints and incomplete information [10–14]. Fortunately, as a bottom-up modeling paradigm, MAS is capable of addressing these limitations through several key properties [15,16].

- *Heterogeneity:* A typical MAS consists of one or more agents operating within a simulated environment according to predetermined rules, which may be derived from the literature, expert knowledge, or empirical data analysis [17]. MAS inherently represents heterogeneous agents, such as investors and banks, each endowed with distinct attributes and strategies, thereby capturing the complexity and diversity observed in real financial markets.
- *Autonomy and adaptability:* In MAS, the effects of other agents are frequently incorporated into the environment. This creates an adaptive dynamic system where agents make independent decisions and continuously adjust their strategies based on feedback [18], thereby accurately reflecting the behavior of real economic entities within evolving financial landscapes.
- *Bounded rationality:* Unlike traditional models, MAS incorporates the concept of bounded rationality, where agents make decisions under cognitive and informational constraints. This allows for a more realistic representation of human decision-making processes.

The practical application of MAS in finance directly tackles the core limitations of traditional models, namely, heterogeneity, lack of adaptability, and unrealistic rationality. First, MAS naturally incorporates diverse agents with varying attributes and strategies, enabling researchers to examine how micro-level differences influence macro-level market dynamics, such as asset pricing and systemic risk. Second, to overcome the limited adaptability of traditional models, MAS agents operate autonomously and respond to new information and changing conditions. This autonomy allows the simulation of dynamic phenomena such as market regime shifts and volatility clustering. Third, MAS mitigates the paradox of rational expectations by modeling agents with bounded rationality and memory. Rather than assuming perfect rationality, agents rely on heuristics and past experiences, supporting more realistic simulations of learning, herding behavior, bubbles, and crashes [19].

While traditional MAS represents a significant advancement, the integration of LLMs further enhances these capabilities, creating a paradigm particularly well suited to the complexities of financial markets. LLMs provide agents with advanced cognitive and decision-making capabilities. Accordingly, agents can process and interpret vast amounts of unstructured text data, such as financial news, corporate filings, and social media sentiment, to inform their actions. This capability allows them to emulate the behavior of real-world traders, who respond to market narratives rather than solely to price signals. Furthermore, by resorting to natural language instructions, LLMs facilitate the assignment of complex and nuanced "personas" to agents (e.g., a value investor analyzing quarterly reports versus a momentum trader monitoring online forums), thereby achieving a higher degree of heterogeneity and behavioral realism. This ability allows LLM-based MAS to more faithfully simulate complex market phenomena. The detailed technical architecture and implementation of LLM-based MAS are presented in the following section.

3 Technical analysis of LLM-based agents

This section provides a technical analysis of financial agents (agents applied in the financial scenario), tracing their evolution from traditional rule-based systems to the current LLM-driven paradigm. It reviews the supporting technologies, presents a taxonomy of foundational models used for agents, and outlines the underlying mechanisms that facilitate the emergence of collective intelligence within MAS. An overview of the framework for LLM-based MAS is shown in Figure 2. The enhanced capabilities of individual agents serve as foundational cognitive nodes. When integrated at the system level via structured communication and role specialization, these isolated capabilities synergize to manifest emergent collective intelligence.

3.1 Review of traditional MAS in finance

Originating from the simulation of human social behavior, MAS traces its roots back to the 1940s cellular automata [20]. A seminal early example is Thomas Schelling's 1971 segregation model [21], which demonstrates how macro-level phenomena can emerge from micro-level decisions and contribute to his receipt of the 2005 Nobel Prize in Economics. In the context of finance, agent evolution has progressed from rule-based MAS to adaptive agents with autonomous learning capabilities, marking a shift from designer-defined behaviors to agent-driven policy learning. Initially, financial MASs were predominantly rule-based. These systems, while transparent and interpretable, have been limited by their reliance on predefined trading logic, which restricts their adaptability to new market scenarios.

The integration of reinforcement learning (RL) and deep learning (DL) has shifted the paradigm of financial agents from rule-driven to data-driven models, representing a fundamental transition from designer-defined behaviors to agent-autonomous policy learning. This synthesis, commonly referred to as deep reinforcement learning (DRL),

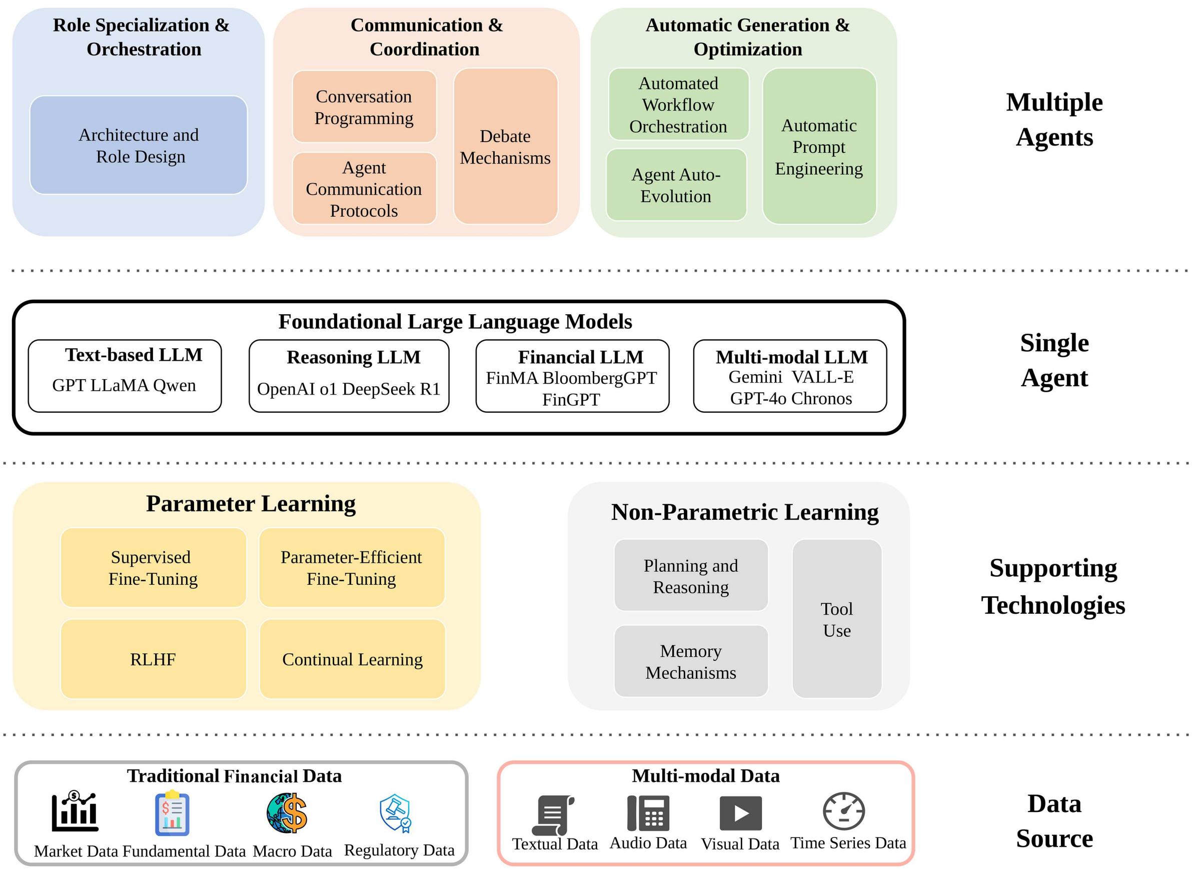

Figure 2 (Color online) An overview of the framework for LLM-based MAS, showing how foundation models (bottom) serve as the intelligent core, powered by rich task context and multi-modal data. This core is then enhanced at the single agent level (middle) through parameter learning techniques and non-parametric learning techniques. These individual agents are subsequently integrated into sophisticated MASs (top), managed through architecture design, role specialization, communication, and automatic optimization to enable collaborative problem-solving.

combines the strengths of both methodologies. Specifically, RL provides a theoretical framework that enables agents to learn optimal strategies through direct interaction with a dynamic environment, such as a financial market. Agents select actions (e.g., buy, sell, hold), receive feedback via a reward signal (e.g., profit, loss, or risk-adjusted return), and iteratively refine their policies to maximize cumulative rewards [22]. DL employs deep neural networks as powerful, universal function approximators. These networks can process large volumes of raw, nonlinear, and non-stationary market data, such as price histories and limit order books. By doing so, they learn the complex representations of value functions or policies necessary for decision-making in such environments [23]. This DRL approach has been successfully applied across various financial domains, such as finance portfolio management [24], market simulation [25], and optimal execution [26].

Despite their adaptability, DRL-based systems face several challenges. The stochastic and non-stationary nature of financial markets can easily lead to overfitting. Additionally, the behaviors learned by agents may not fully capture the bounded rationality of human traders. Together, these factors limit the explanatory power of these models in economic contexts. In recent years, LLM-based agents have emerged as a key technological paradigm in complex domains like computational finance. This emergence is driven by significant advancements in natural language understanding, knowledge transfer, and complex reasoning. This transition significantly enhances agents' cognitive and interactive capabilities within intricate financial environments.

3.2 Supporting technologies of LLM-based agents

The supporting technologies for LLM-based agents can be broadly categorized into two groups: nonparametric methods and parametric methods. These technologies collectively enhance the core capabilities of individual agents, providing a foundation for their application in financial markets.

3.2.1 *Nonparametric level approaches*

Individual LLM-based agents form the foundational building blocks for more complex systems. Their core capabilities originate from the underlying LLM and are further augmented through specialized frameworks, primarily in the following areas.

Planning and reasoning: By employing prompting techniques including chain-of-thought (CoT) [27], ReAct [28], and Reflection [29], LLMs are capable of generating interpretable decision logic trees that emulate step-by-step reasoning mechanisms. To ensure logical rigor, prompts generally follow the RBTO (role-background-task-output) paradigm. Within this paradigm, CoT is frequently utilized for sequential financial statement interpretation. Meanwhile, ReAct establishes a thought-action-observation loop intended to align internal reasoning with real-time market signals. This capability enables agents to partition complex financial tasks into a series of actionable sub-tasks, thereby facilitating structured and systematic decision-making.

Tool use: Real-world financial tasks often require interaction with external data sources and systems. By incorporating modular toolchains (e.g., market data APIs and event semantic parsers) [30–32], agents can establish interactive channels with the environment, enabling end-to-end automation for data retrieval, information extraction, and real-time decision-making. To ensure systematic reliability, these tools are often encapsulated as skills following standardized schemas such as the Anthropic skill specification. Each skill is defined by a distinct name and semantic description for indexing, an input_schema (JSON Schema) for strict parameter constraints, and a unified execution logic to generate structured outputs. Studies such as Toolformer [33] have further demonstrated that LLMs can autonomously learn to utilize external tools, substantially expanding their action space.

Memory mechanisms: Memory modules play a crucial role in maintaining contextual consistency and facilitating learning from experience over extended time horizons. By incorporating retrieval-augmented generation (RAG) techniques [34, 35], agents can dynamically access domain knowledge from structured sources, such as transaction logs, and unstructured texts including research reports. In financial scenarios, RAG systems always employ hybrid chunking to preserve the semantic integrity of tables. Additionally, they use multi-vector indexing to correlate visual charts with textual descriptions. Finally, they apply domain-specific reranking to prioritize documents based on key financial indicators like volatility and net margin. This capability enables agents to align long-term strategic planning with short-term behavioral adaptation, supporting continual learning of market dynamics and risk patterns.

3.2.2 *Parameter level approaches*

In addition to the general-purpose, non-fine-tuning enhancement frameworks described above, fine-tuning models with domain-specific data constitutes another essential pathway for developing high-performance and high-reliability single-agent systems. This approach encodes desired behavioral patterns, domain knowledge, and alignment preferences directly into the model parameter space, resulting in more stable and efficient performance for specialized tasks. Key methodologies include the following.

Supervised fine-tuning (SFT): By training on high-quality and diverse instruction-response datasets, SFT guides the model to adhere to specific formats and styles. SFT serves as a primary mechanism for embedding domain knowledge, including financial terminology, regulatory requirements, and foundational task capabilities such as summarization and question answering. Through this process, the model can initially adopt the behavioral paradigm of a professional assistant within specific contexts, thereby establishing a foundation for more complex alignment.

Parameter-efficient fine-tuning (PEFT): To address the high computational and storage costs of full SFT, PEFT methods such as low-rank adaptation (LoRA) [36] have been developed. LoRA freezes the majority of a pretrained model's weights and introduces small low-rank adapter matrices that are trained alongside key layers. Specifically, for financial task fine-tuning, LoRA is frequently configured with a rank r of 8 or 16 and an alpha α of 32 to balance task-specific plasticity and model stability. These adapters are typically applied to the query and value projection layers within the attention mechanism to capture specialized market nuances without compromising general linguistic capabilities. Moreover, by minimally perturbing the original model parameters, LoRA helps prevent catastrophic forgetting while facilitating the acquisition of new skills [37].

RL from human feedback (RLHF): When success metrics cannot be easily defined through static datasets, RLHF becomes essential for aligning model behavior. It first trains a reward model to learn human preferences over model outputs, and then uses this reward model as a signal within an RL loop to optimize the agent's policy [38, 39]. In financial contexts, this alignment often prioritizes risk-sensitive preference mapping and regulatory compliance by incorporating domain-specific constraints into the reward modeling process. Recent methods, such as direct preference optimization, streamline this process by converting preference data into a stable, single-stage classification objective, thereby avoiding the complexity and instability associated with reward modeling [40].

Continual learning mechanisms: In real-world financial environments, market rules, data patterns, and user requirements are constantly evolving. Continual learning allows agents to incrementally adapt without necessitating full retraining. This is typically achieved by combining PEFT methods with memory systems. For example, a new LoRA adapter can be trained for a specific task or data period and integrated into the primary agent via model merging [41] or dynamic adapter composition. Such mechanisms allow the agent to iteratively update its capabilities, which is crucial for long-term autonomous operation.

3.3 Taxonomy of foundational models for financial agents

With the development of supporting technologies, LLMs are increasingly demonstrating trends toward diversification and specialization. Different categories of LLMs play critical roles in fundamental cognition, cross-modal perception, temporal data processing, and complex reasoning. Furthermore, they serve as foundational modules for perception, reasoning, and execution within financial agent architectures. Based on their functional emphasis and applicable scenarios, the prevailing large models used to construct financial agents can be classified into four main types.

(1) Base LLMs represent the core driving force behind the current wave of AI. Typically built upon autoregressive decoder-only or encoder-decoder architectures, they serve as the default "brain" for most intelligent agents. Representative models include early versions of OpenAI's GPT series [42–44], Meta's LLaMA series [45], and Anthropic's Claude series. These models are pretrained on large-scale corpora of text and code, endowing them with strong capabilities in language understanding, generation, reasoning, and knowledge integration. Designed primarily for general-purpose use, base LLMs can rapidly adapt to a wide range of tasks via in-context learning and instruction tuning, ranging from content creation and code synthesis to conversational interaction. Thus, they provide the fundamental cognitive and operational framework upon which more specialized models and agent architectures are constructed.

(2) Reasoning LLMs are optimized for advanced logical deduction, mathematical reasoning, and complex planning tasks. Leveraging the parameter-level approaches (e.g., RLHF and SFT) discussed in Subsection 3.2.2 on high-quality datasets, these models exhibit clear and interpretable multi-step reasoning chains, such as CoT [27], tree of thoughts (ToT) [46], and program-aided language (PAL) [47]. As a result, they substantially outperform general-purpose models in terms of logical rigor and intermediate step verification. The DeepSeek series exemplifies this category. Trained using large-scale RL and cold-start data, these models generate self-reflective reasoning chains and self-verification processes and are subsequently distilled into multiple scales ranging from 1.5 B to 70 B parameters [48]. Within agent architectures, reasoning LLMs typically function as planners or master controllers, responsible for high-level strategy formulation, task decomposition, and execution auditing. Their reliability and transparency render them highly promising in academic research, engineering, and decision support domains, where they can even surpass human expert performance.

(3) Financial LLMs integrate the general capabilities of LLMs with deep domain-specific expertise, with particular optimization for financial terminology, regulatory compliance, and market dynamics. Although general-purpose LLMs exhibit broad knowledge, they often suffer from hallucinations or insufficient understanding in specialized financial contexts. To address these limitations, industry leaders have developed domain-specific models such as BloombergGPT [49], FinGPT [50], FinMA [51], Cfgpt [52], RA-CFGPT [53], as well as proprietary models by major financial institutions. The emergence of such specialized models has further motivated the design of dedicated evaluation frameworks, such as CFBenchmark for the Chinese financial domain, to assess model performance and identify areas for improvement [54]. By applying the domain-specific fine-tuning techniques from Subsection 3.2.2 to large-scale corpora comprising financial news, research reports, trading data, and regulatory documents, domain-specific prior knowledge is embedded within their parameters. As a result, they represent well-suited foundations for constructing highly reliable financial analysts, risk managers, and investment advisor agents.

(4) Multimodal LLMs significantly enhance the capabilities of financial agents by integrating diverse sensory modalities, including vision, audio, and temporal data, into a unified intelligence framework. Vision-language models, such as GPT-4o [43], Gemini [55], and LLaVA [56], combine advanced language understanding with visual perception, allowing agents to interpret images, charts, and complex financial visualizations like candlestick graphs and scanned reports. Audio-language models, such as Whisper [57] and Vall-E [58], further extend agent interaction into the auditory domain. Consequently, agents can process earnings calls, investor meetings, and financial podcasts through real-time transcription, sentiment analysis, and multilingual understanding. Complementing these modalities, time-series LLMs, such as TimeGPT [59], Chronos [60], and Lag-LLaMA [61], are specifically designed to analyze sequential financial data and to model trends and anomalies across stock prices and macroeconomic indicators. Together, these multimodal models empower agents with comprehensive perceptual capabilities, enabling robust and context-aware decision-making in dynamic financial environments. To systematically summarize the

Table 2 Performance, modality, and cost comparison of representative foundational LLMs for financial agents.

Model name	Base model	Open	MM	MATH-500 (Acc) (%)	GPQA (Dia.) (%)	FinEval (Avg) (%)	FinMR (Acc) (%)	Est. cost (in/out) (USD/1M Tok)
XuanYuan3-70B-Chat	Llama	Yes1)	No	N/A	N/A	59.1	N/A	N/A
FinGPT v3.1	Llama	Yes2)	No	N/A	N/A	27.1	N/A	N/A
FinBERT	BERT	Yes3)	No	N/A	N/A	N/A	N/A	N/A
Llama 4 Maverick	Llama	Yes4)	Yes	90.0	N/A	N/A	27.34	N/A
DeepSeek-R1	DeepSeek	Yes5)	Yes	97.3	71.5	N/A	44.84	N/A
Qwen2.5-72B-Instruct	Qwen	Yes6)	No	83.1	49.0	69.4	N/A	N/A
OpenAI o1	GPT	No	Yes	96.4	75.7	N/A	45.78	$15.00/$60.00
GPT-4o	GPT	No	Yes	74.6	49.9	71.9	50.00	$2.50/$10.00
Claude 3.5 Sonnet	Claude	No	Yes	78.3	65.0	72.9	N/A	$3.00/$15.00
Gemini 2.5 Pro	Gemini	No	Yes	98.0	N/A	N/A	54.76	$1.25/$5.00
Qwen2.5-VL-72B	Qwen	Yes7)	Yes	N/A	N/A	N/A	34.84	N/A
DeepSeek-VL2	DeepSeek	Yes8)	Yes	N/A	N/A	N/A	37.81	N/A
InternVL3-78B	InternLM	Yes9)	Yes	N/A	N/A	N/A	31.87	N/A

1) https://huggingface.co/Duxiaoman-DI/Llama3-XuanYuan3-70B-Chat.
2) https://github.com/AI4Finance-Foundation/FinGPT.
3) https://huggingface.co/ProsusAI/finbert.
4) https://github.com/meta-llama/llama-models.
5) https://github.com/deepseek-ai/DeepSeek-R1.
6) https://huggingface.co/Qwen/Qwen2.5-72B-Instruct.
7) https://huggingface.co/Qwen/Qwen2.5-VL-72B-Instruct.
8) https://github.com/deepseek-ai/DeepSeek-VL2.
9) https://huggingface.co/OpenGVLab/InternVL3-78B.

Table 3 Summary of financial LLM-based MAS and their key characteristics.

Model	Open source	Modality	Key techniques
StockAgent [62]	Yes	Text	Multi-agent interaction, event-driven simulation
TradingGPT [63]	No	Multimodal	Layered memory, inter-agent debate
Alpha-GPT 2.0 [64]	No	Text	Human-in-the-loop, iterative refinement
MLAB [65]	No	Text	Multi-agent LLM mapping by education/cognition level
TradingAgents [66]	Yes	Multimodal	LLM task allocation, collaboration, simulation environment
EconAgent [67]	Yes	Text	Perception module, memory module, action module
EconAI [68]	No	Text	Preference-driven LLM, knowledge brain, real-time data integration
TWINMARKET [69]	Yes	Multimodal	Belief-desire-intention, information aggregation mechanism, order-driven trading system
GenSim [70]	Yes	Multimodal	Distributed parallel computing, error-correction mechanisms
TaxAgent [71]	No	Text	Adaptive tax policy, iterative feedback mechanism
EconGym [72]	Yes	Multimodal	Cross-domain task coordination, modular economic modeling

characteristics of the aforementioned models, Table 2 presents a comparison of their general reasoning performance, modality support, and estimated deployment costs.

3.4 The rise of collective intelligence: LLM-based MAS

Essentially, the financial markets constitute complex ecosystems composed of numerous heterogeneous participants, such as value investors, high-frequency traders, macroeconomic analysts, and retail investors. These participants differ in their information sets, goals, and strategies, and their interactions, cooperation, and strategic competition collectively give rise to market dynamics.

Therefore, in order to more realistically simulate the complexity of financial markets, the shift from a single agent to the "collective intelligence" of an MAS is not only a technological evolution but also a theoretical necessity. For example, during an interest rate shock, a single-agent model predicts a uniform price drop. Conversely, an MAS captures heterogeneous interactions, including the dynamic between high-frequency traders short-selling and value investors buying undervalued assets. This heterogeneous modeling framework thereby generates realistic emergent volatility that single-agent paradigms cannot reproduce. LLM-based MAS is specifically designed to address this challenge by shifting the research paradigm from isolated executors to a simulated "agent society", as summarized in Table 3 [62–72]. Recently, progress in this field has primarily focused on the following core mechanisms.

3.4.1 *System architecture & role specialization*

LLM-based MAS often adopts modular and role-specialized architectures, in which individual agents are assigned distinct identities and professional functions. To effectively coordinate these heterogeneous roles, a variety of architectural strategies have emerged. Among them, two primary paradigms dominate: hierarchical and distributed architectures.

Hierarchical architectures are the most commonly adopted design paradigm. They operate under a top-down control flow, in which a central orchestrator or manager agent decomposes complex tasks and sequentially assigns subtasks to specialized expert agents. MetaGPT [73] exemplifies this approach by conceptualizing the agent system as a software company with roles such as product manager, architect, and engineer. Each agent follows standard operating procedures (SOPs), generates structured documents, and communicates via shared message queues to form a clear production pipeline. Similarly, ChatDev [74] simulates the complete software development lifecycle.

In contrast, distributed architectures adopt a peer-to-peer topology, where agents operate as equals without a centralized coordinator. Consensus and collaboration are achieved through negotiation, communication, or voting. This structure is more flexible and robust, making it particularly suitable for creative, exploratory, or adversarial scenarios such as multi-agent debates [75].

In practice, hybrid architectures that combine both hierarchical and distributed topologies are increasingly adopted. For example, systems like FinAgent [76] and TradingAgent [66] typically employ a multi-layered design. At the top level, a hierarchical structure is used, where a manager agent handles workflow and sequentially invokes data analysis agents. At the middle level, a distributed structure is employed, in which a strategy committee composed of diverse agents (e.g., fundamental analysts, technical analysts, and macroeconomists) engages in peer-level debate and consensus formation. At the bottom level, the architecture returns to a hierarchical mode with a report generation agent consolidating the consensus into a coherent and logically rigorous final output. In addition, some studies in economics model agents as adhering to explicit social norms or game-theoretic rules, where agents simultaneously engage in cooperation and competition, forming structured societies that simulate broader "social" environments. For instance, a sandbox environment has been designed in [77], where generative agents can form friendships, organize gatherings, and adapt their behaviors based on past experiences. Such social structures are particularly suitable for applications in economic simulations and sociological research, where large-scale agent interactions are essential. However, these systems often suffer from limited controllability and reduced predictability.

3.4.2 *Communication & coordination*

Effective inter-agent communication is a fundamental prerequisite for the emergence of collective intelligence. Accordingly, the existing research has explored a range of communication paradigms and coordination mechanisms, including the following.

Conversation programming: A typical work in this direction is the AutoGen framework [78], which models workflows as dialogues among multiple agents. In this framework, agents are conversational and can dynamically adapt their behaviors based on dialogue history and context. The GroupChatManager component enables turn-taking and dynamic speaker selection, facilitating both structured discussions and unstructured brainstorming. Such dialogue-based coordination is highly applicable in finance, where agents with opposing market views (e.g., bullish, bearish, or neutral) debate investment decisions.

Agent communication protocols: Frameworks such as AutoGen are typically limited to homogeneous agent ecosystems. To support cross-platform and cross-organization collaboration, general communication protocols are required. These include context-oriented protocols like Anthropic's model context protocol. This protocol uses standardized JSON-RPC interfaces to provide tools and data as context. Additionally, inter-agent protocols like A2A and ANP establish formal standards for communication and coordination among heterogeneous agents. These communication protocols form the foundational infrastructure for building composable and interoperable systems.

Debate mechanisms: To improve decision robustness, debate-oriented frameworks, such as LLM-Debate [75], ACC-Debate [79], and Chateval [80], encourage agents to present competing solutions, challenge each other, and converge on reliable decisions via voting or aggregation. This is particularly valuable in risk-sensitive applications, such as financial risk assessment.

3.4.3 *Automatic generation & optimization*

The manual design of high-performing multi-agent teams remains a labor-intensive process. Consequently, the automatic generation and optimization of agents and their collaborative processes have become a frontier research area.

Table 4 Comprehensive summary of technical pros and cons specifically for financial scenarios.

Category	Specific technology	Pros	Cons
Traditional MAS	Rule-based	Highly transparent and interpretable trading logic.	Fail to adapt dynamically to novel market scenarios.
	DRL	Autonomous policy learning from complex, non-linear market data.	Overfit in stochastic markets; misses human bounded rationality.
Agent enhancements	Non-parametric	Automate real-time market API execution and correlate long-term strategy context.	High indexing overhead for financial reports; vulnerable to API failures.
	Parametric	Embed strict financial compliance and deep terminology while retaining stability.	High update costs for regime shifts; complex reward modeling.
Foundational LLMs	Base LLMs	Rapid adaptation to diverse tasks via in-context learning.	Prone to hallucinations in specialized financial contexts.
	Reasoning LLMs	High logical rigor for multi-step investment strategy verification.	Heavy computational overhead for generating long reasoning chains.
	Financial LLMs	Deep expertise in regulatory compliance and market dynamics.	Narrow applicability; require dedicated domain evaluation.
	Multimodal LLMs	Unify visual (K-lines), audio (earnings calls), and time-series data.	Complex and resource-intensive integration across diverse inputs.
MAS mechanisms	Hierarchical	Establish clear SOPs for structured workflows.	Top-down control flow may limit flexibility and adaptability.
	Distributed	Capture emergent macro-level dynamics through peer-level debate.	Limited controllability and reduced system predictability.
	Comm. & debate	Improve decision robustness in risk-sensitive applications via cross-validation.	Require standardized cross-platform interoperability protocols.
	Auto-optimization	Dynamically route tasks and evolve structures without manual design.	Extremely high theoretical and computational graph complexity.

Automated workflow orchestration: The automatic orchestration of workflows across diverse environments enhances both the adaptability and collaborative efficiency of MASs. Mainstream approaches focus on abstracting the architecture of agent systems into optimizable computational graphs or dynamic networks, and incorporating search and optimization mechanisms to jointly evolve structure and policy. For example, GPTSwarm [81] is the first to abstract an MAS as an optimizable computational graph, distinguishing between node-level optimization (i.e., agent capabilities) and edge-level optimization (i.e., inter-agent collaboration and information flow). This framework provides a rigorous theoretical foundation for automated system design. Building on this, frameworks such as AFLOW [82] and MaAS [83] represent agent workflows as code and utilize automated search techniques to dynamically optimize the connectivity among nodes. DyLAN [84] abstracts multi-agent collaboration as a temporal feed-forward network and proposes a two-stage process of team optimization followed by task resolution. MasRouter [85] further advances this paradigm by learning how to route different tasks to the most suitable LLMs within the system, co-optimizing both agent roles and collaboration modes. During task execution, it dynamically restructures collaboration patterns and team membership via an agent team reformation mechanism, thereby achieving task-oriented, efficient, and adaptive collaboration.

Agent auto-evolution: Agent auto-evolution aims to replace manual design with automated search and optimization strategies, thereby enhancing the performance of LLM-based agents. For example, AgentVerse [86] introduces the concept of expert recruitment, dynamically forming agent teams based on the task complexity. Similarly, systems such as ADAS [87] and AgentSquare [88] utilize LLMs to build agent generators, enabling the automated design and optimization of agent prompts and structures, thereby providing diversified automated support for agent design.

Automatic prompt engineering: In the field of automatic prompt engineering, representative methods can be grouped into several categories. Techniques such as OPRO [89] leverage LLMs themselves as optimizers, generating a set of candidate prompts per iteration and selecting the most effective one based on task-specific feedback. Subsequent research, including Textgrad [90] and other text-gradient-based methods, treats prompts as differentiable parameters and optimizes them in continuous space using estimated gradients, providing an essential complement to gradient-driven optimization. DSPy [91] extends beyond conventional prompt paradigms by modularizing prompt engineering into compilable language model programs. Its compiler, termed the teleprompter, jointly optimizes prompts, instructions, and examples across modules, reducing redundancy and enabling end-to-end optimization with training data. Finally, evolutionary strategies, such as EvoPrompt [92] and PromptBreeder [93], leverage genetic algorithms to perform natural language-based prompt mutation and recombination, while simultaneously evolving the mutation strategies themselves to achieve dual-level optimization. To conclude this section, Table 4 provides a comprehensive summary of the technical pros and cons of the aforementioned methodologies specifically for financial scenarios.

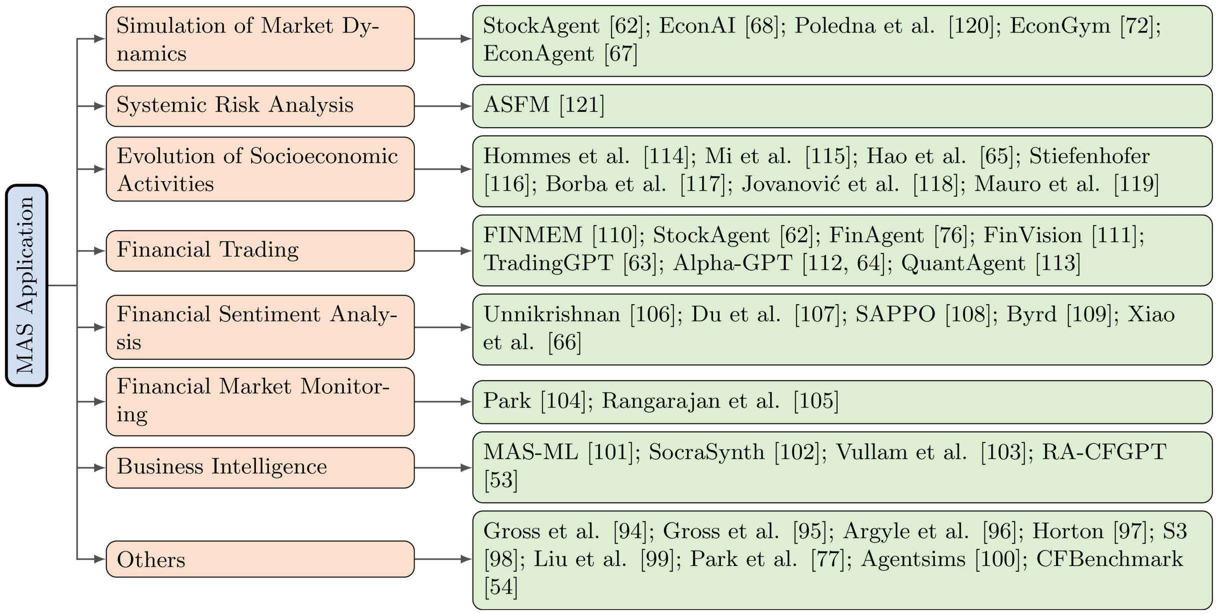

Figure 3 (Color online) An overview of typical studies on the application of LLM-based MAS in financial markets in recent years.

4 LLM-based MAS applications

In this section, we provide a comprehensive overview of the applications of LLM-based MAS in financial markets. Specifically, we examine how they are employed to simulate market dynamics, analyze systemic risk, and model socioeconomic activities. Moreover, we discuss their growing role in tasks such as financial transactions, sentiment analysis, and market monitoring. Representative recent studies on the application of LLM-based MAS in financial markets are presented in Figure 3 [53, 54, 62–68, 72, 76, 77, 94–121].

When deploying LLM-based MAS for simulation, researchers typically follow the workflow outlined below.

(1) Specification of agents and interactions. Identify the types of agents, delineate their action and state spaces, and define protocols governing agent-to-agent interactions.

(2) Implementation and testing. Translate the behavioral rules and interaction protocols into executable code and conduct rigorous unit and integration testing.

(3) Parameter estimation and calibration. Estimate model parameters and calibrate them against stylized empirical facts drawn from real-world data.

(4) Statistical analysis of outputs. Perform batch simulations under varied parameter settings, apply appropriate summary statistics, and use the resulting micro-level data to illuminate the mechanisms that generate observed macro-level patterns.

With the advancement of computing technology and the maturation of behavioral economics theory, the application of LLM-based MAS in financial markets has gradually shifted from theoretical exploration to empirical validation, and from qualitative analysis to quantitative prediction.

4.1 Market dynamics simulation

Early research on MAS in market dynamics simulation focuses on exploring the self-organization of market institutions [122, 123], replicating stylized facts [124], and investigating competitive strategies such as collusion [125] and market power [126]. With the development of economic theory and computational capabilities, MAS has increasingly been integrated with market dynamics models. This integration facilitates stability analysis [127–129] and the examination of institutional structures [130]. Additionally, it drives the creation of large-scale simulation platforms, such as the EURACE project [131–133]. Currently, the primary focus of ABM lies in its predictive capacity, with frameworks increasingly challenging traditional models such as VAR and DSGE in forecasting applications [120, 134, 135].

Recently, research has focused on integrating LLMs into agent architectures to endow them with advanced cognitive abilities. A pioneering work, EconAgent [67], uses LLM-based perception, memory, and reasoning modules to create agents whose economic behaviors more closely resemble human decision-making, moving beyond the static rules of traditional MAS. Empirically, EconAgent successfully replicates key macroeconomic stylized facts, such as the Phillips curve and Okun's law, while maintaining inflation (±5%) and unemployment (2%–12%) within realistic empirical bounds. To address the challenge that LLMs often behave as purely rational entities, subsequent models such as EconAI [68] have built upon the foundation of EconAgent. By incorporating a knowledge brain and self-learning mechanisms, EconAI enables agents to make decisions that balance economic benefits with personal preferences. Consequently, EconAI exhibits superior predictive precision; compared to EconAgent, it reduces the forecasting error for inflation from 0.197 to 0.146, while achieving an unemployment forecasting error of 0.112.

The main application of LLM-based MAS in market dynamics simulation lies in replicating stock market behavior to capture nuanced participant interactions. StockAgent [62] is a representative work in this regard, reflecting recent advancements in general economic agent design. Simulations reveal backbone-dependent heterogeneity: GPT-driven agents manifest aggressive trading behaviors, generating significantly higher trading volumes (14.1 million versus 3.1 million) and exhibiting lower trading frequencies relative to Gemini-driven agents.

The development of these sophisticated LLM-based agents necessitates advanced evaluation platforms. Deploying large-scale agent networks to simulate the entire market in actual business environments faces extremely high computational resource consumption and scalability bottlenecks. To overcome the scalability challenges, researchers have developed test platforms such as EconGym [72]. EconGym provides a scalable testbed supporting up to 10000 heterogeneous agents and diverse economic tasks. Benchmark experiments validate the fidelity of EconGym, demonstrating that the Wasserstein distance consistently decreases as the agent population scale expands. Such a platform facilitates robust benchmarking of AI policies for complex challenges, including demographic shifts and policy coordination, within realistic large-scale simulations.

4.2 Systemic risk analysis

MAS serves as a pivotal tool for analyzing systemic risk. It simulates how micro-level interactions propagate through financial networks and clarifies how systemic risk emerges from the interplay of real-economy shocks and financial contagion, underscoring the need for differentiated policy responses [136]. These models reveal how credit dynamics contribute to economic instability and are used to test macroprudential policies such as countercyclical rules and a Systemic Risk Tax [137, 138]. Recent research without LLM has utilized advanced multilayer and bilateral network models to better measure direct and indirect contagion, assess the impact of monetary policy, and coordinate stability regulations [139–142].

When building a real-time risk monitoring system, it is extremely difficult to seamlessly integrate dynamically evolving corporate fundamentals, diverse real economic behaviors, and broader real financial mechanisms (such as real-time interfaces with banking and lending systems) into the systemic risk analysis. The introduction of LLMs marks a paradigm shift. The system can directly access and parse unstructured interaction data and text reports in the real world.

In contrast to the traditional methods, where interactions are governed by predefined mathematical rules, the interactions between LLM-based agents are driven by reasoning over natural language, as demonstrated in ASFM [121]. This approach enables a more nuanced simulation of market participant behavior, capturing the complex, human-like biases and decision-making processes that are often absent in conventional models. Empirically, ASFM demonstrates robust market engagement through achieving 1855 orders and a 62.26% turnover rate. However, its simulation validity under complex macroeconomic shocks remains constrained by the omission of broader financial mechanisms (e.g., banking and lending), dynamic corporate fundamentals, and diverse real-economy agent behaviors.

4.3 Evolution of socioeconomic activities

As a "computational laboratory", MAS offers unique advantages in studying social phenomena and individual behaviors [143]. Researchers can modify agent strategies and environmental parameters to observe emergent dynamic processes, such as traffic patterns [144], market fluctuations [145], and cultural diffusion [146], which are often overlooked by traditional equilibrium analysis. This approach enables the replication of complex social phenomena from simple rules and the identification of critical conditions for their emergence, making MAS a powerful tool for policy simulation, social science research, and market modeling [147].

Early work in MAS has established its value for analyzing complex economic issues where traditional models fall short. For example, Keynes-Schumpeter frameworks have demonstrated how institutional changes affect inequality and economic stability [148, 149]. Subsequently, the 2008 financial crisis has exposed the limitations of representative-agent DSGE models, prompting a shift toward heterogeneous agent models capable of capturing endogenous crisis dynamics [150]. This foundational period has also witnessed the development of practical policy tools and methods for generating high-fidelity synthetic populations, confirming MAS's ability to replicate real-world statistical patterns.

Recent work has built on this foundation, incorporating greater institutional detail and leveraging AI to address contemporary economic challenges. Models now analyze the granular, distributional impacts of events like pandemics [118] and the potential labor displacement from AGI [116]. A key trend involves employing AI for policy optimization. For instance, RL can be used to design wealth taxes that significantly reduce the Gini coefficient to 0.48 and achieve a peak per capita GDP of 1.7×10^7 [115]. This approach performs significantly better than traditional DSGE models.

Building on these advances, the latest paradigm integrates LLMs to create agents with more realistic behavioral sensitivities [65], which in turn improves the forecasting accuracy. The frontier is moving toward sophisticated digital twins, such as the ECB's EUROABM, which aim to provide real-time, dynamic simulations of entire economies for policy analysis. Deploying such systems faces significant challenges. These include the real-time dynamic calculation of the entire economic system and the need for extremely high-fidelity data alignment. Nevertheless, methods like reinforcement learning (e.g., TaxAI mentioned above) offer a viable solution. They enable rapid dynamic simulation and policy impact evaluation while effectively controlling huge computational costs.

4.4 Financial trading

Agents show high potential in the competitive field of financial trading, using strategic reasoning and adaptability to navigate volatile market conditions [151]. Early research has laid a critical foundation, demonstrating that even simple adaptive agents could approximate human trading behavior [152]. Subsequent frameworks have introduced more sophisticated architectures, incorporating features such as time-sensitive pricing and real-time knowledge exchange to enhance multi-agent coordination [153]. In recent years, the rapid advancement of LLMs has transformed this landscape. These models can process large volumes of financial data, identify patterns, and extract insights. LLM-based agents formulate trading strategies informed by rich contextual information, improving both adaptability and predictive accuracy.

To achieve this, recent research has focused on several key themes: enhancing agents' situational awareness through multimodal data, designing sophisticated cognitive architectures that emulate human reasoning, and building robust LLM-based MAS for collaborative analysis. These efforts collectively aim to move beyond simple automation, enabling agents to strategically navigate complex market dynamics.

One of the most significant trends is the push to create a more holistic understanding of the market by integrating diverse data types. Several recent systems exemplify this shift by moving beyond text to incorporate visual and numerical inputs. For example, FinAgent [76] incorporates a multimodal intelligence module that jointly processes numerical data, financial news, and K-line charts, using a two-stage reflection mechanism to adapt to market shifts. Empirically, FinAgent achieves a 92.27% annual return on TSLA (84.39% relative improvement) and averages over 36% profit improvement across six datasets against 12 baselines including FinMem, PPO, and MACD. Similarly, FinVision [111] integrates multiple specialized LLM-based agents, each focusing on a distinct task: parsing news, analyzing K-line chart patterns, or interpreting trading signals. In addition, the system employs a visual reflection module to enhance its decision-making capacity. In a different domain, a dual-agent framework has been employed in [154] for due diligence in structured finance, enabling the automatic cross-validation of loan applications and bank statements.

Alongside data integration, a parallel effort seeks to imbue agents with more human-like cognitive abilities. Recent research has explored cognitive modeling and memory-enhanced architectures to emulate human trading behavior. For example, FINMEM models human cognition using hierarchical memory to balance short-term signals with long-term strategy [110]. Experiments show FinMem achieves competitive returns (e.g., 44.72% ARR on ETHUSD). However, it reveals unstable decision-making characterized by frequent buy-then-sell reversals and hallucinations regarding cash reserves, leading to buy recommendations despite insufficient funds. Similarly, TradingGPT integrates hierarchical memory with inter-agent debate, enabling collaborative strategic deliberation [63]. QuantAgent uses a nested-loop design in which an inner loop provides tactical responses while an outer loop iteratively refines knowledge through simulation [113]. StockAgent focuses on fundamental analysis, evaluating how macroeconomic indicators and firm-specific data impact investment results within a leakage-free simulation [62].

When deploying LLM-driven high-frequency trading strategies in cloud infrastructures such as serverless environments, issues such as cost, latency, and black-box characteristics are encountered. Currently, some efforts have been made to alleviate the infrastructure-level challenges necessary for practical deployment and credibility. For instance, in [155], the deployment of LLMs has been investigated in serverless environments, highlighting the trade-offs in cost, latency, and scalability. To meet the critical need for transparency, a self-reflective framework has been introduced in [156], which is capable of generating interpretable, natural-language rationales for stock predictions. Finally, recognizing the value of human expertise, Alpha-GPT and its 2.0 iteration [64, 112] adopt a human-in-the-loop paradigm, integrating expert feedback into the α-mining process.

4.5 Financial sentiment analysis

Sentiment analysis, a key task in financial NLP, aims to quantify opinions in text to inform market predictions. Early dictionary-based methods, which analyze news and social media [157], have struggled with limited lexical coverage and semantic context. To overcome these limitations, modern approaches use LLMs to achieve more accurate assessments of market sentiment. This sentiment is then integrated as a qualitative signal into autonomous trading agents, which use RL to refine their strategies based on LLM-parsed news, financial reports, and social media data [66, 106].

LLM-based MAS for financial sentiment analysis draw on diverse textual sources, including news, social media such as Twitter (X), financial forums, and regulatory filings, often integrating them to capture a multi-dimensional view of sentiment. For example, it has been revealed that managerial tone in earnings calls can serve as a predictor of stock returns [107], and some simulated agents equipped with LLMs have learned to manipulate social sentiment for strategic purposes [109]. Leveraging the advanced natural language understanding of LLMs, these agents can interpret domain-specific terminology and synthesize multiple inputs, such as combining positive earnings reports with favorable social media sentiment, to inform trading decisions. Modern systems now apply these sentiment analysis frameworks across nearly all genres of financial text to extract actionable market signals [66, 109].

Empirical evidence has demonstrated that incorporating sentiment enhances the profitability and robustness of trading strategies. For instance, a sentiment-augmented PPO agent (SAPPO) incorporates a sentiment layer extracted from news by an LLM (LLaMA 3.3 [158]). This model achieves higher Sharpe ratios and lower drawdowns than the standard PPO baseline [108,159]. Similarly, other studies have confirmed that RL portfolio managers using LLM-derived sentiment outperform traditional benchmarks [106]. By combining price data with textual signals, these agents can detect market shifts that price-only models often miss, leading to superior trading outcomes.

However, financial sentiment analysis still faces many challenges in real-world deployment. In high-frequency and volatile live trading, the real-time capture and processing of massive multi-dimensional social media and news data place extremely high demands on the system's throughput. Additionally, deployed automated agents are highly vulnerable to maliciously spread false information or market sentiment manipulation attacks with strategic purposes. The industry often uses engineering architectures to intercept 99% of low-value or simple data in cheap and high-speed components, allowing only the core 1% of complex data to occupy expensive LLM inference computing power. Moreover, by double-verifying price data and text signals, the robustness of agents against manipulation and extreme market shifts is enhanced.

In short, a strong convergence emerges between agent-based AI and financial sentiment analysis. By embedding LLMs into trading agents, researchers are augmenting quantitative models with qualitative signals from news and social media. This integration has established agent-based modeling as an increasingly vital approach in financial sentiment research, gaining relevance in parallel with ongoing advances in AI and NLP.

4.6 Financial market monitoring

Financial market surveillance, which is crucial for maintaining market integrity, has traditionally relied on rigid, expert-defined rules that often struggle to keep pace with the speed and complexity of modern trading. These legacy systems tend to be slow, generate high false alarm rates, and prove ineffective against sophisticated algorithmic strategies such as spoofing. In addition, a 2015 CFA Institute survey has confirmed that market manipulation remains a prevalent issue, underscoring the urgent need for more dynamic and intelligent oversight. In response, agent-based technology, particularly when integrated with LLMs, is emerging as a critical evolution for surveillance systems. Current research has focused on several key directions: enhancing anomaly detection through collaborative AI, generating synthetic data to train anti-money laundering models, and using simulations to proactively uncover market manipulation tactics. These approaches collectively shift the paradigm from reactive rule-following toward proactive, adaptive analysis.

To address real-world deployment challenges, including the high latency and high false positive rates of monitoring systems, when deploying models such as anti-money laundering models, they are subject to strict privacy regulations like GDPR. Additionally, financial institutions cannot share real customer proprietary data for joint training. In recent academic explorations, an important trend is the integration of multi-agent collaboration with privacy-preserving technologies. For instance, a multi-agent LLM framework has been developed in [104], where specialized agents collaborate on data analysis, significantly reducing the need for manual review.

At the same time, to combat financial crimes such as money laundering, agent-based simulation has been used to overcome the critical challenge of data scarcity. By generating high-fidelity synthetic datasets, these models provide the necessary ground truth for training and validating machine learning algorithms. For example, an agent-based generator has been proposed in [160] to create realistic transaction datasets with embedded money laundering schemes. These simulations build complex, multi-layered fund flows, offering an innovative and essential tool for benchmarking AML algorithms.

Furthermore, LLM-based MAS serves as a powerful laboratory for detecting and preventing market manipulation. By simulating trader behaviors, researchers can analyze the impact of illicit strategies and test countermeasures. For example, simulations have demonstrated how “probing agents” can degrade dark pool liquidity [105] and how RL agents can be guided away from illegal spoofing by adjusting their reward functions [161].

4.7 Business intelligence and sales strategies

In business intelligence and sales, agent-based technologies are driving a fundamental shift from static decision-support tools to dynamic and autonomous systems. The convergence of LLM-based MAS enables end-to-end solutions that span the entire enterprise workflow, ranging from the optimization of internal operations to the personalization of customer interactions and the automation of strategic planning.

In supply chain management, a primary research focus is on using agent-based systems with advanced machine learning to automate complex operational decisions and improve forecasting accuracy. These frameworks excel at integrating heterogeneous, real-time data to enhance system resilience. For example, the MAS-ML framework has been introduced in [101], where distributed agents apply machine learning to Internet of Things (IoT) data to optimize inventory and logistics, demonstrating particular effectiveness during pandemic-related disruptions. At the customer interface, ongoing research has focused on developing dynamic systems that optimize long-term profitability and user satisfaction by modeling consumer utility and adapting to feedback. Agent-based simulations have shown that incorporating consumer utility yields higher long-term profits than aggressive targeting [162]. Related work has also introduced dynamic supplier coalitions based on customer ratings to improve fulfillment performance [163], and employed clustered MAS to create dynamic user groups, significantly improving recommendation specificity [103].

In the realms of sales and customer engagement, the reasoning capabilities of LLMs are being incorporated into LLM-based MAS to automate strategic planning and interaction processes. LLMs play a crucial role across multiple stages of the customer lifecycle, including awareness and acquisition, consideration and engagement, purchase, and post-purchase (e.g., onboarding and retention) [164]. Rather than merely enhancing operational efficiency, LLMs serve as a foundational technology for elevating customer experience. Their ability to analyze extensive volumes of data and generate insights throughout the customer lifecycle significantly enhances the effectiveness of business processes and customer interactions. Within these systems, agents function as collaborative partners that complement human teams. For instance, a framework named SocraSynth has been introduced in [102], in which multiple LLM-based agents utilize a debate-based approach to formulate effective sales plans and enhance customer engagement. Similarly, in [165], a case study has further illustrated that LLM-based agents, when interpreting real-time market data, can reduce the human workload by 40% while simultaneously improving customer satisfaction.

It should be noted that in actual enterprise operations, systems will inevitably encounter continuous and unpredictable dynamic disruptions, and it is difficult to efficiently integrate and process heterogeneous, real-time IoT data streams from frontline devices. At the same time, secure knowledge sharing within enterprises, as well as the generalization of strategies and the utilization of historical experience, is also a challenge that cannot be ignored. To address the first challenge, a distributed framework (such as the aforementioned MAS-ML) can be adopted. In such frameworks, agents are deployed at the data collection end to directly apply machine learning to IoT data, enabling dynamic inventory and logistics optimization. Regarding the second challenge, current research has actively explored how agents can learn more effectively and share knowledge securely. For instance, a two-stage reflection mechanism has been proposed in [76], which enhances policy generalization by leveraging historical context and offers a pathway for tighter integration between LLMs and RL.

In summary, LLM-based MASs are evolving from back-end analytical tools into proactive collaborators in enterprise decision-making. Their ability to optimize workflows, enhance user experience, and support strategic planning is driving a paradigm shift from experience-driven to cognition-driven business intelligence.

4.8 Others

Beyond the domains discussed above, LLM-based MAS has demonstrated broad influence across many domains due to the advanced language understanding, reasoning, and task-solving capabilities of LLMs. In social science, such agents have been used to simulate human behavior in psychological experiments [166] and provide mental well-being support [167]. In political science and economics, they have been employed for ideology detection, voting pattern prediction [96], and simulating economic behaviors [97]. Moreover, LLM-based MAS enables large-scale social simulations that were previously impractical, thereby facilitating the study of phenomena such as information propagation in virtual communities [98,99] or daily life in simulated towns [77,100]. In law, agents assist in decision-making by simulating judges [168] and acting as specialized legal models that reduce hallucinations [169]. These agents also serve as research assistants, streamlining work by generating abstracts and identifying new research questions [170, 171].

In natural science, agents can support the management of vast scientific literature and data by querying databases and the internet [172, 173]. When deployed as laboratory assistants, they can automate the design, planning, and execution of experiments, offering procedural recommendations and safety warnings [172, 173]. In education, they have been used as specialized tutors across domains such as science [173], mathematics [174], and programming [175], thereby providing personalized support and automated feedback [176]. While these applications demonstrate the versatility of LLM-based MAS, their effectiveness across diverse domains remains constrained by domain-specific hallucination risks and the lack of standardized cross-disciplinary evaluation benchmarks.

5 Limitations and future trends

This section summarizes the primary challenges and future directions of LLM-based MAS in finance. Specifically, we first outline the key limitations, such as data quality, model interpretability, market uncertainty, and regulatory concerns. Then, we present several emerging trends, including continuous adaptation, enhanced human-machine collaboration, and the automated design of large-scale simulations.

5.1 Limitations

5.1.1 *Technical limitations*

Complex deep learning models often reduce transparency in exchange for higher accuracy. Additionally, newer LLMs remain vulnerable to hallucinations, which lead to the generation of inaccurate or fictitious content. To mitigate these risks in live deployments, financial institutions frequently pair complex LLMs with strict, rule-based guardrails for output verification. Financial markets are non-stationary and experience regime shifts, making historical data an unreliable indicator of future trends. Agents also struggle with environmental uncertainty caused by market noise and model uncertainty, both of which amplify model risks during highly volatile periods. Furthermore, benchmarking LLM-based MAS introduces unique open challenges. Even with platforms like EconGym or CFBenchmark, comprehensive evaluation is critically hindered by reproducibility issues stemming from LLM stochasticity, high evaluation sensitivity to prompt design, and the severe scalability constraints of simulating massive, concurrent agent interactions.

5.1.2 *Practical limitations*

The effectiveness of financial agents relies on training data that is frequently difficult and expensive to procure. Furthermore, financial data are often fragmented and may harbor biases or omissions, yielding inaccurate or discriminatory model outcomes. Designing these systems involves a fundamental trade-off between the complexity of the model and the necessity for low-latency, real-time responsiveness.

5.1.3 *Regulatory limitations*

Severe privacy regulations, such as the GDPR, combined with financial institutions' reluctance to share proprietary data, construct substantial barriers to gathering sufficient training samples. Using biased data to train financial agents can magnify societal biases, highlighting the need for strong governance and fairness-by-design. Moreover, the complexity of assigning accountability for AI-driven decisions has prompted regulators to classify these financial systems as high-risk entities needing strict oversight. The lack of unified AI regulations across different jurisdictions introduces significant operational and compliance risks for global financial institutions.

5.2 Future trends

5.2.1 *Continuous learning and adaptation*

Financial markets are unpredictable, and historical data quickly become outdated. To navigate this uncertainty, future models will focus on continuous learning to autonomously evolve through environmental interaction [177–179]. To bypass the high costs of full retraining and avoid catastrophic forgetting (where new data overwrites existing knowledge), systems will increasingly utilize efficient techniques like PEFT and RAG to dynamically absorb the latest market information. Furthermore, to address strict data privacy constraints across financial institutions, federated learning will emerge as a critical methodology, enabling multi-agent systems to continuously adapt using decentralized proprietary data without exposing sensitive information [180–182]. To overcome the technical benchmarking hurdles associated with continuous adaptation, future systems need to integrate standardized, dynamic evaluation protocols. By isolating LLM stochasticity and leveraging automated prompt engineering, these frameworks can ensure consistent reproducibility and robust performance assessment even as agents continuously evolve.

5.2.2 *Human-machine collaboration*

Deep learning models function as opaque black boxes, conflicting with regulatory demands for transparency and complicating the assignment of accountability. To mitigate these transparency gaps, future frameworks will deeply embed explainable AI through concrete methodologies such as self-reflective rationale generation and feature attribution to create more trustworthy systems [183–185]. To mitigate hallucination risks in live deployments, financial institutions will increasingly pair complex LLMs with strict, rule-based guardrails for output verification. A human-machine hybrid intelligence model will emerge, relying on AI to provide data-driven suggestions while empowering humans to make final strategic judgments, thereby establishing a clearer framework for accountability and risk management [186–188]. Rather than a transitional solution, this human-in-the-loop design is anticipated to remain a long-term architectural necessity. It ensures that while AI handles high-throughput data processing, human experts retain ultimate strategic oversight to satisfy strict regulatory and accountability standards.

5.2.3 *Automated agent design and large-scale simulation*

Manually configuring complex agent systems requires extensive expertise, and agents are often constrained by the limited availability of high-quality, real-world data. Emerging methodologies will automate the design and optimization of agent teams and their collaborative workflows [189–191]. To achieve this, methodologies such as reinforcement learning from AI feedback (RLAIF) and evolutionary algorithms will be crucial for autonomously searching and optimizing agent architectures without human intervention. This automation enables the creation of massive digital twins of entire economies, such as the European Central Bank's EUROABM, that bypass fragmented historical data to provide dynamic, real-time analysis for robust policy evaluation [192–195].

Overall, a clear roadmap for future work must explicitly address two overarching research gaps: the scalability bottleneck of simulating massive, high-frequency agent networks, and the critical absence of standardized cross-disciplinary evaluation metrics. To resolve the scalability issue, researchers should increasingly leverage macro-level architectural solutions, such as modularized agent workflows, scalable cloud computing infrastructures, and serverless deployments, to dynamically allocate resources and decouple complex tasks. Concurrently, to address the evaluation gap, developing unified, modular benchmarking environments that isolate agent reasoning from environmental noise will be essential for rigorous and reproducible cross-domain validation.

6 Conclusion

In this paper, a comprehensive survey has been provided on the applications of LLM-based MAS in financial markets, which offers a superior, bottom-up approach compared to traditional models. LLM-based MAS has enabled more dynamic and granular simulations by modeling the behavior of autonomous, heterogeneous, and adaptive agents. The recent integration of LLMs has begun to transform the field, equipping agents with advanced cognitive and reasoning abilities for more realistic and nuanced financial simulations. LLM-based MAS has proven useful in a wide range of financial applications, including simulating stock market dynamics, analyzing systemic risk, and modeling socioeconomic activities such as income distribution. It has also been increasingly applied in financial trading, market monitoring, and anti-money laundering efforts. Nevertheless, significant challenges have remained, including data quality, model interpretability, market uncertainty, and ethical and regulatory concerns.

Addressing these challenges has been crucial for the advancement of LLM-based MAS. Future directions will focus on developing continuous learning capabilities for LLM-based agents, fostering human-machine collaboration, and advancing research in areas such as multimodal learning and efficient model deployment. In summary, LLM-based MAS represents a transformative approach to understanding and navigating the complexities of the global financial landscape, offering powerful tools for research in financial theory, policy making, and risk management.

Acknowledgements This work was supported in part by National Key Research and Development Program of China (Grant No. 2022YFB450-1704), National Natural Science Foundation of China (Grant No. 62473285), Shanghai Science and Technology Innovation Action Plan Project of China (Grant No. 22511100700), Fundamental Research Funds for the Central Universities of China, Royal Society of the U.K., and Alexander von Humboldt Foundation of Germany.

References

1 Farmer J D, Foley D. The economy needs agent-based modelling. Nature, 2009, 460: 685–686
2 Nie Y, Kong Y, Dong X, et al. A survey of large language models for financial applications: progress, prospects and challenges. 2024. ArXiv:2406.11903
3 Axtell R L, Farmer J D. Agent-based modeling in economics and finance: past, present, and future. J Economic Literature, 2025, 63: 197–287
4 Zhao H, Liu Z, Wu Z, et al. Revolutionizing finance with LLMs: an overview of applications and insights. 2024. ArXiv:2401.11641
5 Dong M M, Stratopoulos T C, Wang V X. A scoping review of ChatGPT research in accounting and finance. Int J Accounting Inf Syst, 2024, 55: 100715
6 Li Y, Wang S, Ding H, et al. Large language models in finance: a survey. In: Proceedings of the Fourth ACM International Conference on AI in Finance, 2023. 374–382
7 Hill R, Myatt A. Overemphasis on perfectly competitive markets in microeconomics principles textbooks. J Economic Education, 2007, 38: 58–76
8 Walras L. Elements of Pure Economics: Or the Theory of Social Wealth. London: Routledge, 2013
9 Münnix M C, Shimada T, Schäfer R, et al. Identifying states of a financial market. Sci Rep, 2012, 2: 644
10 Muth J F. Rational expectations and the theory of price movements. Econometrica, 1961, 29: 315–335
11 Becker G S, Murphy K M. A theory of rational addiction. J Political Economy, 1988, 96: 675–700
12 Poole W, Phelps E S, Baily M N. Rational expectations in the macro model. Brookings Papers on Economic Activity, 1976, 1976: 463–514
13 Haltiwanger J, Waldman M. Rational expectations and the limits of rationality: an analysis of heterogeneity. American Economic Rev, 1985, 75: 326–340
14 Simon H A. Theories of bounded rationality. Decision Organization, 1972, 1: 161–176
15 Wooldridge M, Jennings N R. Intelligent agents: theory and practice. Knowledge Eng Rev, 1995, 10: 115–152
16 Macal C M, North M J. Tutorial on agent-based modeling and simulation. In: Proceedings of the Winter Simulation Conference, 2005. 14
17 Crooks A T, Heppenstall A J. Introduction to agent-based modelling. In: Proceedings of Agent-Based Models of Geographical Systems, 2011. 85–105
18 Dosi G, Roventini A, Russo E. Endogenous growth and global divergence in a multi-country agent-based model. J Economic Dyn Control, 2019, 101: 101–129
19 Zhang Z, Dai Q, Bo X, et al. A survey on the memory mechanism of large language model based agents. ACM Trans Inform Syst, 2025, 43: 155
20 Bonabeau E. Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci USA, 2002, 99: 7280–7287
21 Schelling T C. Dynamic models of segregation. J Math Sociology, 1971, 1: 143–186
22 Bai Y, Gao Y, Wan R, et al. A review of reinforcement learning in financial applications. Annu Rev Stat Its Application, 2025, 12: 209–232
23 Mienye E, Jere N, Obaido G, et al. Deep learning in finance: a survey of applications and techniques. AI, 2024, 5: 2066–2091
24 Hu Y J, Lin S J. Deep reinforcement learning for optimizing finance portfolio management. In: Proceedings of AICAI, 2019. 14–20
25 Maeda I, deGraw D, Kitano M, et al. Deep reinforcement learning in agent based financial market simulation. J Risk Financ Manag, 2020, 13: 71
26 Fang Y, Ren K, Liu W, et al. Universal trading for order execution with Oracle policy distillation. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2021. 107–115
27 Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. In: Proceedings of Advances in Neural Information Processing Systems, 2022. 35: 24824–24837
28 Yao S, Zhao J, Yu D, et al. ReAct: synergizing reasoning and acting in language models. In: Proceedings of ICLR, 2023
29 Shinn N, Cassano F, Gopinath A, et al. Reflexion: language agents with verbal reinforcement learning. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 36: 8634–8652
30 Ruan J, Chen Y, Zhang B, et al. TPTU: task planning and tool usage of large language model-based AI agents. In: Proceedings of NeurIPS 2023 Foundation Models for Decision Making Workshop, 2023

31 Kong Y, Ruan J, Chen Y, et al. TPTU-v2: boosting task planning and tool usage of large language model-based agents in real-world industry systems. In: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track, Miami, 2024. 371–385
32 Zhang G, Yue Y, Li Z, et al. Cut the crap: an economical communication pipeline for LLM-based multi-agent systems. 2024. ArXiv:2410.02506
33 Schick T, Dwivedi-Yu J, Dessì R, et al. Toolformer: language models can teach themselves to use tools. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 36: 68539–68551
34 Gao Y, Xiong Y, Gao X, et al. Retrieval-augmented generation for large language models: a survey. 2024. ArXiv:2312.10997
35 Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Proceedings of Advances in Neural Information Processing Systems, 2020. 33: 9459–9474
36 Hu E J, Shen Y, Wallis P, et al. LoRA: low-rank adaptation of large language models. In: Proceedings of ICLR, 2022
37 Biderman D, Portes J, Ortiz J J G, et al. LoRA learns less and forgets less. 2024. ArXiv:2405.09673
38 Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. In: Proceedings of Advances in Neural Information Processing Systems, 2022. 35: 27730–27744
39 Stiennon N, Ouyang L, Wu J, et al. Learning to summarize with human feedback. In: Proceedings of Advances in Neural Information Processing Systems, 2020. 33: 3008–3021
40 Rafailov R, Sharma A, Mitchell E, et al. Direct preference optimization: your language model is secretly a reward model. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 36: 53728–53741
41 Yang E, Shen L, Guo G, et al. Model merging in LLMs, MLLMs, and beyond: methods, theories, applications and opportunities. 2024. ArXiv:2408.07666
42 Achiam J, Adler S, Agarwal S, et al. GPT-4 technical report. 2023. ArXiv:2303.08774
43 Hurst A, Lerer A, Goucher A P, et al. GPT-4o system card. 2024. ArXiv:2410.21276
44 Jaech A, Kalai A, Lerer A, et al. OpenAI o1 system card. 2024. ArXiv:2412.16720
45 Touvron H, Lavril T, Izacard G, et al. LLaMA: open and efficient foundation language models. 2023. ArXiv:2302.13971
46 Yao S, Yu D, Zhao J, et al. Tree of thoughts: deliberate problem solving with large language models. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 36: 11809–11822
47 Gao L, Madaan A, Zhou S, et al. PAL: program-aided language models. In: Proceedings of International Conference on Machine Learning, 2023. 10764–10799
48 Guo D, Yang D, Zhang H, et al. DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning. 2025. ArXiv:2501.12948
49 Wu S, Irsoy O, Lu S, et al. BloombergGPT: a large language model for finance. 2023. ArXiv:2303.17564
50 Liu X Y, Wang G, Yang H, et al. FinGPT: democratizing internet-scale data for financial large language models. 2023. ArXiv:2307.10485
51 Xie Q, Han W, Zhang X, et al. Pixiu: a large language model, instruction data and evaluation benchmark for finance. In: Proceedings of the 37th International Conference on Neural Information Processing Systems, 2023. 33469–33484
52 Li J, Bian Y, Wang G, et al. CFGPT: Chinese financial assistant with large language model. 2023. ArXiv:2309.10654
53 Li J, Lei Y, Bian Y, et al. RA-CFGPT: Chinese financial assistant with retrieval-augmented large language model. Front Comput Sci, 2024, 18: 185350
54 Lei Y, Li J, Cheng D, et al. CFBenchmark: Chinese financial assistant benchmark for large language model. 2023. ArXiv:2311.05812
55 Team G, Anil R, Borgeaud S, et al. Gemini: a family of highly capable multimodal models. 2023. ArXiv:2312.11805
56 Liu H, Li C, Wu Q, et al. Visual instruction tuning. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 36: 34892–34916
57 Radford A, Kim J W, Xu T, et al. Robust speech recognition via large-scale weak supervision. In: Proceedings of International Conference on Machine Learning, 2023. 28492–28518
58 Chen S, Wang C, Wu Y, et al. Neural codec language models are zero-shot text to speech synthesizers. IEEE Trans Audio Speech Language Process, 2025, 33: 705–718
59 Garza A, Challu C, Mergenthaler-Canseco M. TimeGPT-1. 2023. ArXiv:2310.03589
60 Ansari A F, Stella L, Turkmen A C, et al. Chronos: learning the language of time series. 2024. arXiv:2403.07815
61 Rasul K, Ashok A, Williams A R, et al. Lag-Llama: towards foundation models for time series forecasting. In: Proceedings of Robustness of Few-shot and Zero-shot Learning in Large Foundation Models, 2023
62 Zhang C, Liu X, Zhang Z, et al. When AI meets finance (StockAgent): large language model-based stock trading in simulated real-world environments. 2024. ArXiv:2407.18957
63 Li Y, Yu Y, Li H, et al. TradingGPT: multi-agent system with layered memory and distinct characters for enhanced financial trading performance. 2023. ArXiv:2309.03736
64 Yuan H, Wang S, Guo J. Alpha-GPT 2.0: human-in-the-loop AI for quantitative investment. 2024. ArXiv:2402.09746
65 Hao Y, Xie D. A multi-LLM-agent-based framework for economic and public policy analysis. 2025. ArXiv:2502.16879
66 Xiao Y, Sun E, Luo D, et al. TradingAgents: multi-agents LLM financial trading framework. 2024. ArXiv:2412.20138
67 Li N, Gao C, Li M, et al. EconAgent: large language model-empowered agents for simulating macroeconomic activities. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, 2024. 15523–15536

68 Cao Z, Liu A, Xie J, et al. EconAI: preference-driven agents simulating economic activities via large language model. 2025. https://openreview.net/forum?id=HzG3A0VD1k

69 Yang Y, Zhang Y, Wu M, et al. TwinMarket: a scalable behavioral and social simulation for financial markets. 2025. ArXiv:2502.01506

70 Tang J, Gao H, Pan X, et al. GenSim: a general social simulation platform with large language model based agents. 2024. ArXiv:2410.04360

71 Lin J, Sun L, Yan Y. Simulating macroeconomic expectations using LLM agents. 2025. ArXiv:2505.17648

72 Mi Q, Yang Q, Fan Z, et al. EconGym: a scalable AI testbed with diverse economic tasks. 2025. ArXiv:2506.12110

73 Hong S, Zhuge M, Chen J, et al. MetaGPT: meta programming for a multi-agent collaborative framework. In: Proceedings of the Twelfth International Conference on Learning Representations, 2023

74 Qian C, Liu W, Liu H, et al. ChatDev: communicative agents for software development. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2024. 15174–15186

75 Du Y, Li S, Torralba A, et al. Improving factuality and reasoning in language models through multiagent debate. In: Proceedings of the Forty-First International Conference on Machine Learning, 2023

76 Zhang W, Zhao L, Xia H, et al. FinAgent: a multimodal foundation agent for financial trading: tool-augmented, diversified, and generalist. 2024. ArXiv:2402.18485

77 Park J S, O'Brien J, Cai C J, et al. Generative agents: interactive simulacra of human behavior. In: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 2023. 1–22

78 Wu Q, Bansal G, Zhang J, et al. AutoGen: enabling next-gen LLM applications via multi-agent conversations. In: Proceedings of the First Conference on Language Modeling, 2024

79 Estornell A, Ton J F, Yao Y, et al. ACC-Collab: an actor-critic approach to multi-agent LLM collaboration. In: Proceedings of the Thirteenth International Conference on Learning Representations, 2025

80 Chan C M, Chen W, Su Y, et al. ChatEval: towards better LLM-based evaluators through multi-agent debate. In: Proceedings of the Twelfth International Conference on Learning Representations, 2024

81 Zhuge M, Wang W, Kirsch L, et al. GPTSwarm: language agents as optimizable graphs. In: Proceedings of the Forty-First International Conference on Machine Learning, 2024

82 Zhang J, Xiang J, Yu Z, et al. AFlow: automating agentic workflow generation. 2024. ArXiv:2410.10762

83 Zhang G, Niu L, Fang J, et al. Multi-agent architecture search via agentic supernet. 2025. ArXiv:2502.04180

84 Liu Z, Zhang Y, Li P, et al. A dynamic LLM-powered agent network for task-oriented agent collaboration. In: Proceedings of the First Conference on Language Modeling, 2024

85 Yue Y, Zhang G, Liu B, et al. MasRouter: learning to route LLMs for multi-agent systems. 2025. ArXiv:2502.11133

86 Chen W, Su Y, Zuo J, et al. AgentVerse: facilitating multi-agent collaboration and exploring emergent behaviors. In: Proceedings of the Twelfth International Conference on Learning Representations, 2024

87 Hu S, Lu C, Clune J. Automated design of agentic systems. 2024. ArXiv:2408.08435

88 Shang Y, Li Y, Zhao K, et al. AgentSquare: automatic LLM agent search in modular design space. 2024. ArXiv:2410.06153

89 Yang C, Wang X, Lu Y, et al. Large language models as optimizers. In: Proceedings of the Twelfth International Conference on Learning Representations, 2023

90 Yuksekgonul M, Bianchi F, Boen J, et al. Optimizing generative AI by backpropagating language model feedback. Nature, 2025, 639: 609–616

91 Khattab O, Singhvi A, Maheshwari P, et al. DSPy: compiling declarative language model calls into self-improving pipelines. 2023. ArXiv:2310.03714

92 Chen A, Dohan D, So D. EvoPrompting: language models for code-level neural architecture search. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 36: 7787–7817

93 Fernando C, Banarse D, Michalewski H, et al. Promptbreeder: self-referential self-improvement via prompt evolution. In: Proceedings of the 41st International Conference on Machine Learning, 2024. 13481–13544

94 Gross M, Siebenbrunner C. Money creation and liquid funding needs are compatible. In: Proceedings of Central Banking, Monetary Policy and the Future of Money, 2022. 154–186

95 Gross M M, Letizia E. To demand or not to demand: on quantifying the future appetite for CBDC. Technical Report, International Monetary Fund, 2023

96 Argyle L P, Busby E C, Fulda N, et al. Out of one, many: using language models to simulate human samples. Polit Anal, 2023, 31: 337–351

97 Horton J J. Large language models as simulated economic agents: what can we learn from homo silicus? Technical report, National Bureau of Economic Research, 2023

98 Gao C, Lan X, Lu Z, et al. S3: social-network simulation system with large language model-empowered agents. 2023. ArXiv:2307.14984

99 Liu H, Sferrazza C, Abbeel P. Chain of hindsight aligns language models with feedback. In: Proceedings of the Twelfth International Conference on Learning Representations, 2024

100 Lin J, Zhao H, Zhang A, et al. AgentSims: an open-source sandbox for large language model evaluation. 2023. ArXiv:2308.04026

101 Farazi M Z R. Enhancing supply chain resilience with multi-agent systems and machine learning: a framework for adaptive decision-making. American J Eng Technol, 2025, 7: 6–20

102 Tsao W K. Multi-agent reasoning with large language models for effective corporate planning. In: Proceedings of 2023 International Conference on Computational Science and Computational Intelligence (CSCI), 2023. 365–370
103 Vullam N, Vellela S S, Reddy V, et al. Multi-agent personalized recommendation system in e-commerce based on user. In: Proceedings of 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2023. 1194–1199
104 Park T. Enhancing anomaly detection in financial markets with an LLM-based multi-agent framework. 2024. ArXiv:2403.19735
105 Rangarajan S, Ventre C. Impact of pinging in financial markets: an agent based study. In: Proceedings of the 17th International Conference on Agents and Artificial Intelligence, 2025. 172–183
106 Unnikrishnan A. Financial news-driven LLM reinforcement learning for portfolio management. 2024. ArXiv:2411.11059
107 Du K, Xing F, Mao R, et al. Financial sentiment analysis: techniques and applications. ACM Comput Surv, 2024, 56: 1–42
108 Kirtac K, Germano G. Leveraging LLM-based sentiment analysis for portfolio allocation with proximal policy optimization. In: Proceedings of ICLR 2025 Workshop on Machine Learning Multiscale Processes, 2025
109 Byrd D. Exploring sentiment manipulation by LLM-enabled intelligent trading agents. 2025. ArXiv:2502.16343
110 Yu Y, Li H, Chen Z, et al. FinMem: a performance-enhanced LLM trading agent with layered memory and character design. IEEE Trans Big Data, 2025, 11: 3443–3459
111 Fatemi S, Hu Y. FinVision: a multi-agent framework for stock market prediction. In: Proceedings of the 5th ACM International Conference on AI in Finance, 2024. 582–590
112 Wang S, Yuan H, Zhou L, et al. Alpha-GPT: human-AI interactive alpha mining for quantitative investment. In: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2025. 196–206
113 Wang S, Yuan H, Ni L M, et al. QuantAgent: seeking holy grail in trading by self-improving large language model. 2024. ArXiv:2402.03755
114 Hommes C, He M, Poledna S, et al. CANVAS: a Canadian behavioral agent-based model for monetary policy. J Economic Dyn Control, 2025, 172: 104986
115 Mi Q, Xia S, Song Y, et al. TaxAI: a dynamic economic simulator and benchmark for multi-agent reinforcement learning. In: Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, 2024. 1390–1399
116 Stiefenhofer P. Artificial general intelligence and the end of human employment: the need to renegotiate the social contract. 2025. ArXiv:2502.07050
117 Borba J S, Gonçalves S, Anteneodo C. Inequality in a model of capitalist economy. Phys A-Stat Mech Its Appl, 2025, 664: 130457
118 Jovanović B, Landesmann M, Reiter O, et al. Structural change, income distribution and unemployment related to COVID-19: an agent-based model. Technical Report, The Vienna Institute for International Economic Studies, 2023
119 Mauro G, Pedreschi N, Lambiotte R, et al. Dynamic models of gentrification. 2024. ArXiv:2410.18004
120 Poledna S, Miess M G, Hommes C, et al. Economic forecasting with an agent-based model. Eur Economic Rev, 2023, 151: 104306
121 Gao S, Wen Y, Zhu M, et al. Simulating financial market via large language model based agents. 2024. ArXiv:2406.19966
122 Vriend N J. Self-organization of markets: an example of a computational approach. Comput Econ, 1995, 8: 205–231
123 Howitt P, Clower R. The emergence of economic organization. J Economic Behav Organiz, 2000, 41: 55–84
124 LeBaron B. Empirical regularities from interacting long- and short-memory investors in an agent-based stock market. IEEE Trans Evol Computat, 2002, 5: 442–455
125 Kimbrough S O, Murphy F H. Strategic bidding of offer curves: an agent-based approach to exploring supply curve equilibria. Eur J Operational Res, 2013, 229: 165–178
126 Nicolaisen J, Petrov V, Tesfatsion L. Market power and efficiency in a computational electricity market with discriminatory double-auction pricing. IEEE Trans Evol Computat, 2002, 5: 504–523
127 Gintis H. The dynamics of general equilibrium. Economic J, 2007, 117: 1280–1309
128 Mérîo B, Borsos A, Hosszú Z, et al. A high resolution agent-based model of the Hungarian housing market. Technical Report, MNB Working Papers, 2022
129 Gilbert N, Hawksworth J C, Swinney P A. An agent-based model of the English housing market. In: Proceedings of AAAI Spring Symposium: Technosocial Predictive Analytics, 2009. 30–35
130 Farmer J D, Patelli P, Zovko I I. The predictive power of zero intelligence in financial markets. Proc Natl Acad Sci USA, 2005, 102: 2254–2259
131 Deissenberg C, van der Hoog S, Dawid H. EURACE: a massively parallel agent-based model of the European economy. Appl Math Computation, 2008, 204: 541–552
132 Dawid H, Gemkow S, Harting P, et al. The eurace@unibi model: an agent-based macroeconomic model for economic policy analysis. SSRN Electron J, 2012, doi: 10.2139/SSRN.2408969
133 Dawid H, Harting P, van der Hoog S, et al. Macroeconomics with heterogeneous agent models: fostering transparency, reproducibility and replication. J Evol Econ, 2019, 29: 467–538
134 Ge J. Endogenous rise and collapse of housing price. Comput Environ Urban Syst, 2017, 62: 182–198
135 Magliocca N, McConnell V, Walls M. Exploring sprawl: results from an economic agent-based model of land and housing markets. Ecol Economics, 2015, 113: 114–125
136 Assenza T, Delli Gatti D, Grazzini J. Emergent dynamics of a macroeconomic agent based model with capital and credit. J Economic Dyn Control, 2015, 50: 5–28
137 Farmer J, Kleinnijenhuis A M, Nahai-Williamson P, et al. Foundations of system-wide financial stress testing with heterogeneous insti-

tutions. Technical Report, Bank of England, 2020
138 Poledna S, Thurner S. Elimination of systemic risk in financial networks by means of a systemic risk transaction tax. Quantitative Finance, 2016, 16: 1599–1613
139 Ding Z, Yan H, Chen Y, et al. Risk contagion in interbank lending networks: a multi-agent-based modeling and simulation perspective. Expert Syst Appl, 2024, 256: 124847
140 Riccetti L. Agent-based multi-layer network simulations for financial systemic risk measurement: a proposal for future developments. Int J Microsimulation, 2022, 15: 44–61
141 Gao Q, Fan H, Pang C. Monetary policy and systemic risk in a financial network system based on multi-agent modeling. Mathematics, 2025, 13: 378
142 Pallante G, Guerini M, Napoletano M, et al. Robust-less-fragile: tackling systemic risk and financial contagion in a macro agent-based model. J Financial Stability, 2025, 76: 101352
143 Dosi G, Roventini A. More is different ... and complex! The case for agent-based macroeconomics. J Evol Econ, 2019, 29: 1–37
144 Axhausen K W, Horni A, Nagel K. The Multi-Agent Transport Simulation MATSim. London: Ubiquity Press, 2016
145 Amrouni S, Moulin A, Vann J, et al. ABIDES-Gym: gym environments for multi-agent discrete event simulation and application to financial markets. In: Proceedings of the Second ACM International Conference on AI in Finance, 2021. 1–9
146 Desmarchelier B, Fang E S. National culture and innovation diffusion. Exploratory insights from agent-based modeling. Tech Forecasting Soc Change, 2016, 105: 121–128
147 Luke S, Balan G C, Panait L, et al. MASON: a Java multi-agent simulation library. In: Proceedings of Agent 2003 Conference on Challenges in Social Simulation, 2003
148 Dosi G, Pereira M C, Roventini A, et al. The effects of labour market reforms upon unemployment and income inequalities: an agent-based model. Socio-Economic Rev, 2018, 16: 687–720
149 Dosi G, Fagiolo G, Napoletano M, et al. Income distribution, credit and fiscal policies in an agent-based Keynesian model. J Economic Dyn Control, 2013, 37: 1598–1625
150 Poledna S, Molina-Borboa J L, Martínez-Jaramillo S, et al. The multi-layer network nature of systemic risk and its implications for the costs of financial crises. J Financial Stability, 2015, 20: 70–81
151 Ding H, Li Y, Wang J, et al. Large language model agent in financial trading: a survey. 2024. ArXiv:2408.06361
152 Cliff D, Bruten J. Less than human: simple adaptive trading agents for CDA markets. IFAC Proc Volumes, 1998, 31: 117–122
153 Luo Y, Liu K, Davis D N. A multi-agent decision support system for stock trading. IEEE Network, 2002, 16: 20–27
154 Wan X, Deng H, Zou K, et al. Enhancing the efficiency and accuracy of underlying asset reviews in structured finance: the application of multi-agent framework. 2024. ArXiv:2405.04294
155 Kathiriya S, Devineni S, Challa N. Serverless architecture in LLMs: transforming the financial industry's AI landscape. Int J Sci Res, 2023, 12: 2131–2136
156 Koa K J, Ma Y, Ng R, et al. Learning to generate explainable stock predictions using self-reflective large language models. In: Proceedings of the ACM Web Conference 2024, 2024. 4304–4315
157 Consoli S, Barbaglia L, Manzan S. Fine-grained, aspect-based sentiment analysis on economic and financial lexicon. Knowledge-Based Syst, 2022, 247: 108781
158 Grattafiori A, Dubey A, Jauhri A, et al. The llama 3 herd of models. 2024. ArXiv:2407.21783
159 Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms. 2017. ArXiv:1707.06347
160 Altman E, Blanuša J, Von Niederhäusern L, et al. Realistic synthetic financial transactions for anti-money laundering models. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 36: 29851–29874
161 Byrd D. Learning not to spoof. In: Proceedings of the Third ACM International Conference on AI in Finance, 2022. 139–147
162 Ghanem N, Leitner S, Jannach D. Balancing consumer and business value of recommender systems: a simulation-based analysis. Electron Commerce Res Appl, 2022, 55: 101195
163 Nanda P, Patnaik S. A multi-agent coalition-based approach for order fulfilment in e-commerce. Decision Analytics J, 2023, 7: 100227
164 Soni V. Large language models for enhancing customer lifecycle management. J Empirical Social Sci Studies, 2023, 7: 67–89
165 George J G. Transforming banking in the digital age: the strategic integration of large language models and multi-cloud environments. Int J Comput Trends Technol, 2024, 72: 77–86
166 Aher G V, Arriaga R I, Kalai A T. Using large language models to simulate multiple humans and replicate human subject studies. In: Proceedings of International Conference on Machine Learning, 2023. 337–371
167 Ma Z, Mei Y, Su Z. Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. In: Proceedings of AMIA Annual Symposium, 2024. 1105–1114
168 Hamilton S. Blind judgement: agent-based supreme court modelling with GPT. In: Proceedings of the AAAI-23 Workshop on Creative AI Across Modalities, 2023
169 Cui J, Li Z, Yan Y, et al. ChatLaw: open-source legal large language model with integrated external knowledge bases. 2023. ArXiv:2306.16092
170 Ziems C, Held W, Shaikh O, et al. Can large language models transform computational social science? Comput Linguist, 2024, 50: 237–291
171 Bail C A. Can generative AI improve social science? Proc Nat Acade Sci, 2024, 121: e2314021121

172 Bran A M, Cox S, Schilter O, et al. Augmenting large language models with chemistry tools. Nat Mach Intell, 2024, 6: 525–535
173 Boiko D A, MacKnight R, Gomes G. Emergent autonomous scientific research capabilities of large language models. 2023. ArXiv:2304.05332
174 Drori I, Zhang S, Shuttleworth R, et al. A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level. Proc Natl Acad Sci USA, 2022, 119: e2123433119
175 Liffiton M, Sheese B E, Savelka J, et al. CodeHelp: using large language models with guardrails for scalable support in programming classes. In: Proceedings of the 23rd Koli Calling International Conference on Computing Education Research, 2023. 1–11
176 Dan Y, Lei Z, Gu Y, et al. EduChat: a large-scale language model-based chatbot system for intelligent education. 2023. ArXiv:2308.02773
177 Liu X, Ling Z, Zhang Y. Torus-event-based sliding mode control for networked interval type-2 fuzzy systems under deception attacks. Int J Network Dynamics Intell, 2025, 4: 100024
178 Li L, Du Y, Liu Y H, et al. Research on image compression encoding based on fixed dictionary. Syst Sci Control Eng, 2025, 13: 2437160
179 Chen L, Wu P, Tan W, et al. A novel UAV-based road damage detection algorithm with lightweight convolution and attention mechanism. Int J Network Dynamics Intell, 2025, 4: 100025
180 Chen W, Hu J, Wu Z, et al. A survey on fault detection for networked systems under communication constraints. Syst Sci Control Eng, 2025, 13: 2460434
181 Guan C, Zheng L, Wang C. A novel CEEMD-based multichannel denoising autoencoder for noise attenuation of surface microseismic data. Int J Network Dynamics Intell, 2025, 4: 100026
182 Sun L, An W, Chen Y, et al. An overview of distributed economic dispatch of microgrids: advances and challenges. Syst Sci Control Eng, 2025, 13: 2467077
183 Ma L, Zhang H, Wang G, et al. Security coordination control for the belt conveyor systems with false data injection attacks. Int J Network Dynamics Intell, 2025, 4: 100027
184 Zhan Y, Yang R, You J, et al. A systematic literature review on incomplete multimodal learning: techniques and challenges. Syst Sci Control Eng, 2025, 13: 2467083
185 Zhang Y, Zhang X, Miao D, et al. Real-time semantic segmentation of road scenes via hybrid dilated grouping network. Int J Network Dynamics Intell, 2025, 4: 100029
186 Zou L, Song B, Suo J, et al. A survey on outlier-resistant state estimation and its applications. Syst Sci Control Eng, 2025, 13: 2474471
187 Yuan S, Ma L, Gao C. Probability-guaranteed consensus control for nonlinear multi-agent systems under bit flips. Int J Network Dynamics Intell, 2025, 4: 100020
188 Song B, Zhao S, Dang L, et al. A survey on learning from data with label noise via deep neural networks. Syst Sci Control Eng, 2025, 13: 2488120
189 Yu L, Ding J, Peng H, et al. Sampled-data based containment control for a class of nonlinear multiagent systems with dynamic leaders and control saturation. Int J Network Dynamics Intell, 2025, 4: 100011
190 Hasan A, Kuncara I, Widyotriatmo A, et al. Secure state estimation and control for autonomous ships under cyberattacks. Syst Sci Control Eng, 2025, 13: 2518964
191 Lan Q, Kaul A, Jones S. Prompt injection detection in LLM integrated applications. Int J Network Dynamics Intell, 2025, 4: 100013
192 Amin A A, Mubarak A, Manzoor H U. Design of intelligent vehicular and sensor communication network: a comprehensive survey. Syst Sci Control Eng, 2025, 13: 2529187
193 Wei Y, Wang Y, Zhu B, et al. Underwater detection: a brief survey and a new multitask dataset. Int J Network Dynamics Intell, 2024, 3: 100025
194 Mawanza J T, Manganda M, Ijiga O E. A novel fixed-time control scheme for multi-UAV wildfire tracking with dynamic reconfiguration and fault tolerance. Syst Sci Control Eng, 2025, 13: 2587840
195 Freek B, Alireza J, Bas O, et al. Analysing the effect of a dynamic physical environment network on the travel dynamics of forcibly displaced persons in Mali. Int J Network Dynamics Intell, 2024, 3: 100003