International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

Multi-Agent AI Systems in Finance: Models, Applications, and Challenges

Dr. Sanjay Nakharu Prasad Kumar

IEEE Senior Member, USA

ABSTRACT: Multi-agent artificial intelligence (AI) systems are emerging as a transformative paradigm in finance, leveraging multiple interactive agents to address complex financial problems that exceed the capabilities of single-model approaches. This article provides a comprehensive examination of multi-agent AI models and their applications in financial decision-making, market simulations, and algorithmic trading. Drawing on insights from Chen and Takamura's *Agent AI for Finance* (2025) and recent empirical research, we analyse how advances in large language models (LLMs) enable agents to collaborate through natural language communication and simulate human-like decision processes. We present evidence that multi-agent systems achieve up to 42% better accuracy in complex financial forecasting tasks [1] and 35% improvement in decision quality compared to single-agent approaches [1], with advantages in scenarios requiring diverse expertise integration.

Key applications examined include multi-agent discussion frameworks for financial data annotation and report generation, hierarchical agent teams that emulate traditional trading desk structures, and agent-based market simulations that capture nuanced participant behaviours. We explore the critical role of LLMs in enabling agent reasoning and communication, introducing the concept of dynamic interaction loops where LLMs and specialized models iteratively refine outputs. The integration of forward-looking financial argument mining enables agents to systematically analyse future-oriented statements and ground decisions in robust scenario analysis, improving risk-adjusted returns by up to 31% in back testing scenarios.

Our analysis reveals significant strengths of multi-agent systems, including enhanced transparency through traceable decision chains (with 87% of compliance officers preferring these systems for regulatory adherence) [3], superior stability (65% lower variance in returns during volatile markets) [2], and modular flexibility allowing dynamic adaptation to changing market conditions. However, we identify substantial challenges including computational costs (3-5x higher than single-agent systems), coordination complexity (23% of failures attributed to coordination breakdowns) [4], potential bias amplification (up to 3.7x stronger than individual agent biases) [3], and security vulnerabilities that concern 67% of financial institutions. [4]

Current adoption remains limited, with 78% of major financial institutions having exploratory projects but only 12% progressing beyond proof-of-concept stages. The field currently exists in what researchers term the "second generation" of multi-agent financial AI, with fully autonomous trading systems projected to emerge in 3-7 years pending solutions to technical, regulatory, and trust barriers. We propose a graduated autonomy approach for deployment, emphasizing the need for standardized evaluation frameworks, robust security architectures, and regulatory frameworks specific to multi-agent decision-making.

This article adds to the expanding literature on AI in finance by presenting a critical review of contemporary capabilities and limitations, providing practitioners and researchers with a guide to building reliable, efficient, and transparent agent-based financial AI systems. While multi-agent AI holds much promise for augmenting financial decision-making through collective intelligence, actualizing this promise will depend on resolving serious technical challenges as well as aligning with financial goals and ethical considerations.

KEYWORDS: Multi-Agent Artificial Intelligence, Large Language Models, Financial Decision-Making, Algorithmic Trading, Forward-Looking Argument Mining

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

I. INTRODUCTION

Finance decision-making frequently entails intricate reasoning, varied sources of information, and coordination among various experts or decision-makers. Conventional AI techniques in finance, e.g., supervised learning algorithms for sentiment analysis or stock forecasting, are normally applied to specific narrow tasks in a siloed manner. Multi-agent AI systems, on the other hand, imagine AI "agents" capable of working together, communicating, and allocating tasks between themselves, similar to a group of analysts or traders collaborating. This method follows from a transition from seeing AI as a tool to seeing AI as an active participant in problem-solving. Recent advances in large language models (LLMs) such as GPT-3.5 and GPT-4 have hastened interest in multi-agent systems, since the models have general reasoning capacity and natural language communication capabilities that enable them to take on varied roles and dynamically interact.

The move towards multi-agent systems is a paradigm shift in the way AI can cope with financial complexity. As demonstrated by recent studies, multi-agent systems can utilize distributed intelligence to handle issues that would be intractable using single-model methods (Singh, 2024). Distributed processing is especially useful in finance, where decisions need to consider market behavior, regulatory requirements, risk exposures, and strategic goals simultaneously. The capacity of the agents to specialize in various areas of financial analysis while keeping coordination through advanced communication protocols provides a more subtle and holistic method for financial decision-making.

Multi-agent AI has the potential to revolutionize the way we produce insights and make decisions in the finance field. Agent AI for finance, according to Chen and Takamura (2025), is a multimodal, interacting multiple agent/model system that takes multimodal inputs (text, audio, images, etc.) and produces useful information to enable human-centred financial decisions. This multimodal ability is essential in contemporary finance analysis, as emphasized by Zhang et al. (2025), who illustrate how multi-agent systems can efficiently handle disparate data streams such as market data, news sentiment, social media signals, and regulatory filings to produce stronger investment insights. For instance, an agent may examine the text of an earnings call, agents may evaluate various aspects like company fundamentals or market sentiment, and another agent may integrate their results into a report on investment. By delegating subtasks to expert agents and then combining their results, such a system simulates a group of expert financial analysts collaborating.

The Agent AI paradigm builds on previous ideas about agent-based modelling in economics – which previously relied on hand-crafted rules or equations to emulate interactions – by leveraging AI agents (notably LLMs) to emulate behaviour in a more natural, language-based way. This innovation has especially profound implications for algorithmic trading and market simulation. According to Al-Kindi (2025), multi-agent systems in algorithmic trading make it possible to design more adaptive and clever trading strategies by enabling agents to specialize in different aspects of the market like technical analysis, fundamental analysis, and risk management, while a coordinating agent aggregates these views into rational trading decisions. This offers potential to model intricate financial situations (e.g. how various market agents respond to a central bank rate increase) using interactive agent conversations instead of just mathematical equations.

The application of multi-agent systems to finance also deals with essential issues of transparency and explainability. Franzen (2024) highlights how multi-agent architectures have the potential to make AI-driven financial decisions more interpretable by making explicit the reasoning process using agent communication logs and decision hierarchies. Transparency is critical for regulatory compliance and trust building with stakeholders that must know how AI systems produce recommendations.

This paper presents a comprehensive overview of multi-agent AI systems in finance, drawing primarily on the content of Agent AI for Finance (Chen & Takamura, 2025) and related research. We first outline the general concepts of multi-agent AI and how LLM-based agent's function. We then explore key applications in finance: collaborative financial decision-making and analysis, agent-based simulations of markets or behaviours, and multi-agent approaches to trading. We discuss specific examples such as multi-agent discussion frameworks that improve financial data annotation and report generation, and hierarchical agent organizations that emulate a traditional trading desk with analysts, traders, and managers. In examining these cases, we highlight the role of LLMs as both individual agents and orchestrators of agent teams.

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

We also introduce the concept of dynamic interaction loops, where LLMs and smaller models iteratively refine each other's outputs, and discuss how this technique can enhance agent performance in tasks like refining financial documents. This iterative refinement process, as demonstrated in recent implementations, can significantly improve the quality of financial analysis by allowing agents to challenge assumptions, verify calculations, and enhance the comprehensiveness of their collective output (Zhang et al., 2025). Additionally, we consider forward-looking financial argument mining – an emerging NLP approach to analyse forward-looking statements in financial text – and its relevance for agent-based systems that need to reason about future scenarios and investment arguments.

Finally, we provide a critical analysis of the current state of multi-agent AI in finance. While early studies demonstrate improved outcomes from agent collaboration (e.g. more accurate annotations or more consistent trading decisions), there are significant challenges. The computational complexity of coordinating multiple agents, as highlighted by Singh (2024), can lead to scalability issues, particularly in high-frequency trading environments where milliseconds matter. We discuss strengths of multi-agent approaches, such as incorporating diverse expertise and more human-like decision processes, as well as limitations including high computational costs, coordination complexity, potential propagation of errors or biases among agents, and difficulties in evaluation of agent-generated content. The paper concludes with a summary and reflections on the path forward for integrating multi-agent AI into real-world financial applications, where reliability and transparency will be as critical as raw performance.

II. OVERVIEW OF MULTI-AGENT AI SYSTEMS

In general terms, a multi-agent AI system consists of multiple intelligent entities (agents) that interact with each other and their environment to solve tasks. Each agent can be an AI model specialized for certain subtasks, and agents may communicate through a common language (in the case of LLM-based agents, via natural language prompts and responses). The definition of an AI agent varies, but a useful description is "a class of interactive systems that can perceive inputs and produce meaningful actions". In the context of finance, agents might perceive data such as market prices, financial news text, or earnings call audio, and produce outputs like analytic summaries or trading decisions.

Recent research has expanded our understanding of how these agents can be effectively organized. Singh (2024) demonstrates that multi-agent systems in finance benefit from adaptive coordination mechanisms that allow agents to dynamically adjust their interaction patterns based on market conditions. This adaptive capability is particularly important in volatile financial markets where rigid organizational structures may fail to respond adequately to rapid changes. Multi-agent systems can be organized in different architectures – from flat structures where all agents are peers discussing on equal footing, to hierarchical structures where agents have roles with different authority levels (e.g. junior analysts vs. senior managers). The choice of architecture depends on the problem and can affect outcomes, as we will see in financial trading scenarios.

The architectural diversity of multi-agent systems offers significant advantages for financial applications. As noted by Al-Kindi (2025), hierarchical multi-agent structures in algorithmic trading can mirror traditional trading desk organizations, with specialized agents for market analysis, risk assessment, and execution coordination. This organizational flexibility allows systems to be tailored to specific financial domains – for instance, a flat structure might be optimal for collaborative research and analysis tasks, while a hierarchical structure could better suit trading operations where clear decision-making chains are essential.

Large language models (LLMs) have become a cornerstone of recent multi-agent frameworks. Because LLMs are trained on vast amounts of text, they can generate coherent language and reason about a wide range of topics. This makes them well-suited as agents that need to handle unstructured financial information (like news articles or reports) and communicate findings. Zhang et al. (2025) emphasize that the multimodal capabilities of modern LLMs enable agents to process not just text but also numerical data, charts, and even audio from earnings calls, creating a more comprehensive analytical framework. Crucially, LLMs can be prompted to adopt a persona or role, enabling one model to act as, say, a "credit risk analyst" and another as a "market strategist."

The advent of ChatGPT and similar LLMs around 2022–2023 spurred the multi-agent trend by showing that multiple LLM agents working together can sometimes outperform a single model working alone. For example, recent studies outside the finance domain found that having multiple LLM-based agents collaborate on a coding task (with roles like manager, coder, reviewer) led to better solutions than a single agent coding in isolation. This evidence has prompted researchers to explore multi-agent setups in finance as well, where complex problems might benefit from a team-of-

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

agents approach. Franzen (2024) provides empirical evidence that multi-agent financial analysis systems can improve decision accuracy by up to 35% compared to single-agent approaches, particularly in complex scenarios requiring diverse expertise such as merger and acquisition evaluations or comprehensive risk assessments. We are essentially moving toward AI systems that mirror human teamwork, in which each agent contributes different skills or perspectives, and they iteratively refine their collective output.

It is important to note that multi-agent AI for finance builds on decades of work in agent-based modelling (ABM) in economics and finance. ABM traditionally involves simulating many autonomous agents (often simple rule-based actors) to study emergent phenomena like market dynamics or contagion. Indeed, ABM has been advocated to capture complex economic systems beyond what equation-based models can do (Farmer & Foley, 2009). Chen and Takamura highlight that agent-based modelling has long been used to simulate real-world interactions in finance, but until recently this was done with simplified agents governed by equations or heuristics.

The evolution from traditional ABM to LLM-based multi-agent systems represents a paradigm shift in financial modelling. As Singh (2024) notes, traditional agent-based models were limited by their reliance on predetermined rules and mathematical formulations, which often failed to capture the nuanced, context-dependent decision-making of real market participants. Now, with advanced AI, we can create agents that behave in a more realistic manner – for instance, by letting LLMs play different roles and interact via natural language, rather than using fixed mathematical rules. This approach can produce simulations that are richer in narrative and potentially closer to how actual market participants think and communicate. For example, one could simulate an entire trading floor or an online investor community by assigning an LLM to each "participant" and observing their dialogue in response to an event (like a sudden interest rate change). While such simulations are still experimental, they illustrate the new horizon that multi-agent AI opens complex adaptive systems in finance can be modelled with agents that have cognitive and communicative abilities.

The practical implementation of these systems requires sophisticated coordination mechanisms. Zhang et al. (2025) describe a novel approach where agents employ "dynamic consensus protocols" – allowing them to reach decisions through iterative negotiation rather than simple voting or averaging. This mirrors how human financial teams operate, with discussion, debate, and eventual consensus-building. Furthermore, their research shows that such consensus mechanisms can help identify and mitigate individual agent biases, leading to more robust collective decisions.

Despite the excitement, it's worth clarifying that multi-agent AI in finance is in an early stage of development. The field lacks standardized definitions and evaluation metrics. For instance, what exactly qualifies as an "AI agent" can vary – some definitions require physical embodiment or the ability to take physical actions, but in financial contexts we often use the term for purely software agents that take information-processing actions. Chen and Takamura acknowledge that the definition of agent-based AI is still evolving, and their working definition emphasizes multi-agent interaction and multimodal input/output in service of human decision-making.

Franzen (2024) proposes a comprehensive evaluation framework for multi-agent financial systems that includes metrics for accuracy, explainability, computational efficiency, and regulatory compliance. This framework addresses a critical gap in the field by providing standardized benchmarks that can facilitate comparison across different multi-agent implementations. As research progresses, it will be important to refine these concepts and develop common frameworks. One concept introduced by Chen and Takamura is the "AI as Partner" paradigm, where AI agents are seen not just as tools to automate tasks but as collaborators that engage in multi-step reasoning and dialogue.

Enabling this collaboration often requires an administrative agent or coordinator that can break complex problems into sub-tasks and assign them to specialized agents, then integrate their results. Al-Kindi (2025) demonstrates that advanced orchestration mechanisms can significantly improve system performance by optimizing task allocation based on each agent's demonstrated expertise and current workload. This kind of orchestration mechanism is a key component in many multi-agent systems (sometimes called a "controller" or "manager" agent). It ensures that the agents' efforts are aligned and that the problem-solving process is organized – analogous to a project manager orchestrating a team of humans.

In summary, multi-agent AI systems bring together multiple AI models to work cooperatively. The rise of LLMs has been a game-changer, endowing agents with versatile language and reasoning capabilities, and enabling natural interaction between agents in scenarios that closely resemble human teamwork. The integration of adaptive coordination mechanisms, multimodal processing capabilities, and sophisticated consensus protocols has created

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

systems that can tackle financial problems with unprecedented sophistication. The next sections delve into how these multi-agent approaches are applied in finance, from decision support and analysis generation to trading simulations.

III. APPLICATIONS IN FINANCIAL DECISION-MAKING AND ANALYSIS

One of the most promising applications of multi-agent AI in finance is enhancing decision-making processes and analytical workflows. Financial analysis often benefits from considering multiple viewpoints or data sources – something a single model might struggle with. By assigning different agents to different aspects of an analysis, we can replicate a "meeting of experts" who discuss and cross-verify information. Chen and Takamura term this a **multi-round discussion** framework. In a multi-round debate, multiple AI agents (e.g., multiple instances of GPT-4 or other models) give their analysis or opinion on a task in the first round. If their responses conflict, the agents proceed to a second round where they compare their reasoning with each other and perhaps revise their conclusions. This can repeat for several rounds, during which a consensus is established (or a vote is held if there is still disagreement). This type of procedure mimics the way that human analysts might consult to label data or write a report, and it has been tried on tasks such as report writing and financial text classification.

Multi-agent discussion for data annotation: In one study, three LLM agents (GPT-4 and others) were used to label financial documents with sentiment or other categories. Each agent annotated the data and explained its reasoning. When disagreements occurred, the agents exchanged explanations and attempted to resolve inconsistencies in a subsequent round. The result was a small but consistent improvement in labelling accuracy – for instance, the multi-agent framework improved annotation accuracy on a financial dataset by about 3.9% compared to a single-agent approach. This suggests that **simulating a team of annotators** can yield higher-quality labels, likely because agents catch each other's errors and pool their knowledge. However, this accuracy gain comes at the cost of more computation (multiple model calls instead of one).

Multi-agent discussion for analysis report generation: Beyond labelling data, multi-agent systems can generate textual analyses that incorporate diverse perspectives. A compelling use case is **financial report writing**. Instead of asking one model to summarize market events, we can have multiple agents *discuss* the events and contribute different insights, then combine them into a report. Chen and Takamura describe an experiment where after a company's earnings call, several LLM agents are tasked with examining different aspects of the situation – for example, one agent reviews the audio transcript for tone and sentiment, another looks at the company's recent financial performance data, and another reads relevant news. Each agent generates a short analysis from its perspective. Then a "writer" agent (which could be another LLM) takes all these inputs and composes a comprehensive earnings analysis report.

However, multi-agent report generation also surfaces challenges. First, evaluating the quality of the report is difficult. The task is inherently subjective – there is no single "correct" analysis, and judging insightfulness or correctness requires human expertise. The researchers tried using LLMs themselves as judges to rank the reports but found that LLMs tended to Favor content produced by AI (often rating AI-generated reports as better than human-written ones). This bias of LLM evaluators indicates a limitation in current automatic evaluation methods. Second, the multi-agent approach did not completely avoid problems like *hallucination* or factual errors – agents might still generate incorrect statements, and if one agent states a falsehood, others could accept and build on it in the discussion. Ensuring factual accuracy and consistency remains an open problem in AI-generated analysis. Nonetheless, the experiment demonstrates a qualitative strength of multi-agent systems: by having agents play different expert roles (much like an equity analyst, industry analyst, etc.), the final output can cover a wider range of considerations and **mimic a reasoned discussion** rather than a unilateral summary.

Hierarchical decision-making and trading: Another application is using multi-agent systems to **replicate decision processes in a structured organization**, such as a financial trading desk or investment committee. In real-world trading firms, decisions often pass through multiple layers – analysts research and recommend trades, traders execute and manage positions, and senior managers oversee risk and approve major decisions. To model this, Chen and Takamura implemented a hierarchical multi-agent framework for trading decisions. In their setup, one agent is assigned the role of **analyst** who, given a piece of news (say a surprise earnings announcement), must write an analysis of the news. A second agent acts as a **trader**, making a trading decision (e.g. "buy" or "sell" a stock) based on that analysis. A third agent plays the **head trader (manager)** who reviews the analysis and the trade decision and decides whether to approve the trade. This simulates a common hierarchical workflow on Wall Street. They compared three approaches: (1) a single agent directly reading the news and making a trading decision; (2) a *flat* multi-agent discussion where

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

multiple agents discuss the news and vote on a decision; and (3) the *hierarchical* approach with distinct analyst/trader/head roles (multi-agent with roles).

The hierarchical multi-agent system performed the best in their study at aligning with real professional traders' decisions. Table 4.1 in Chen & Takamura's book shows that the hierarchical approach had a higher consistency with actual trader decisions and with the subsequent market movement, outperforming both the single-agent and flat discussion approaches. Specifically, the role-based agents were about 2.1% more consistent with professional decisions than a single agent, whereas a multi-agent discussion without hierarchy only gave ~0.2% improvement. This suggests that structuring agents in a realistic organizational hierarchy provides a tangible benefit, likely because it enforces a logical flow of information (analysis → decision → approval) like human processes. Each agent can focus on its role: the analyst-agent digs into why a trade might be good, the trader-agent focuses on how to act on that information, and the head-agent provides a sanity check. This division of labour and oversight seems to lead to decisions that "closely resemble those of professionals".

The hierarchical simulation also revealed some nuanced insights. The authors noticed that prompts encouraging a **long-term investment perspective** led to better agent decisions. In practice, experienced traders think beyond short-term noise; similarly, when the AI agents were prompted to consider longer horizons (as opposed to quick speculative gains), their decisions became more consistent and aligned with expert behaviour. This ties into the notion of *expert knowledge and financial arguments* – real analysts often justify trades with forward-looking theses (e.g. "we expect revenue to grow over the next year due to X, therefore we buy now"). If the AI agents are guided to emulate that style of reasoning, they perform more robustly. It underscores that **prompt engineering** (i.e. how we instruct each agent) is critical: giving agents context about professional norms or longer-term thinking can significantly influence the quality of their output. Another observation was a **seniority bias**: the head trader agent was more likely to approve a decision if the proposing trader agent was described as "senior" rather than "junior," even when the content of the proposal was identical. This is a fascinating emergent behaviour – it mirrors human organizational bias where ideas from senior staff might be given more weight. But in an AI system, this is essentially a flaw, as ideally decisions should be judged on their merit, not on who (which agent) proposed them. The head agent's bias toward a senior title highlights the need for mechanisms to ensure agents focus on rationales and evidence (the analysis) rather than superficial cues. Chen and Takamura relate this to the importance of **financial argument mining** – basically, if agents do not properly consider the underlying arguments and justifications, they can be swayed by irrelevant factors like perceived authority. In designing multi-agent systems, we may need to explicitly correct for such biases, perhaps by anonymizing inputs or by training the agents to evaluate arguments based on content alone.

Finally, the researchers experimented with **cross-agent verification** in decision-making, akin to a double-check system. For example, after the trader agent decided, they tried using another agent to verify the decision (like recalculating or reasoning it out independently). Interestingly, if the verifying agent had the same or lower capability as the original decision agent, it did not help – sometimes it even *hurt* performance. Only when a clearly superior model/agent was assigned to verify the work of a lesser agent did accuracy improve (and even then, it did not exceed simply having the superior agent do the task to begin with). This suggests that naive redundancy (just having two agents do the same thing) isn't very beneficial unless the second agent brings additional skills or knowledge. In other words, **multi-agent systems gain from complementary abilities, not mere duplication**. An agent good at making decisions might not be the best at scrutinizing decisions; verification may require a different skill set. This insight is valuable for system design: we should leverage heterogeneity (different strengths across agents) rather than homogeneity when assigning roles.

In summary, multi-agent AI systems can enhance financial decision-making and analysis by partitioning tasks among specialized agents and mimicking collaborative workflows. Discussion-style frameworks improve upon single-model performance in tasks like annotation and report generation by enabling peer review and diverse inputs (at the cost of more computation). Hierarchical agent teams can reproduce structured decision pipelines found in trading firms, yielding decisions that better align with human experts and market outcomes. These case studies illustrate that **structure and interaction design** in multi-agent systems (e.g. how agents are organized and prompted) are crucial to their success. As we move forward, refining these interactions – ensuring agents share relevant information, heed each other's arguments, and remain unbiased – will be key to unlocking the full potential of multi-agent AI in finance.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

IV. AGENT-BASED SIMULATIONS AND MARKET MODELLING

Beyond aiding discrete decision tasks, multi-agent systems offer a novel approach to **simulations** in finance. Simulation here refers to modelling the behaviour of a complex system (like a financial market or an economy) by recreating the micro-level interactions that drive macro-outcomes. Traditional agent-based models in economics often involve many simple agents (e.g., traders following heuristic rules) to investigate phenomena such as bubbles, crashes, or the impact of regulatory policies. With the introduction of LLM-based agents, we can strive for more **realistic and dynamic simulations**, where agents behave like real human participants, complete with conversational interactions and adaptive strategies.

Chen and Takamura discuss the idea of using LLMs to simulate societal behaviours and note that it is an *early stage yet promising* direction. A striking example outside pure finance is their adaptation of **Schelling's segregation model** using AI agents. Schelling's classic model had agents on a grid who move based on simple preferences, showing how mild individual bias can lead to major segregation. In the AI-augmented version, the "agents" are LLMs making move decisions given prompts about their neighbour's demographics. The experiment introduced LLM-agent suggestions into each simulated individual's decision-making (for instance, an AI agent might advise an individual whether to relocate to a different neighbourhood based on some criteria). The outcome measured how different LLMs (GPT-4, GPT-3.5, etc.) influenced the final segregation level, effectively testing if AI advice could mitigate or exacerbate bias-driven behaviour in aggregate. They found that all the tested LLMs led to an *increase* in segregation compared to the baseline, though the degree varied (with one model causing a 19% increase and others around 25–30%). This indicates that if agents collectively follow certain flawed or biased suggestions, the macro-outcome can be worse than the original scenario. While this simulation isn't about financial markets per se, it offers a cautionary parallel: **if many AI agents (or human decision-makers influenced by AI) operate in a market, their collective biases could amplify systemic risks**. For instance, imagine many trading agents using the same LLM-based strategy – they might all overreact to certain news in the same way, increasing volatility.

In more directly financial contexts, one could use multi-agent systems to simulate **market ecosystems**. For example, different agent models could represent retail investors, institutional traders, market makers, regulators, etc., each with their own goals and information, and then simulate how they interact (by trading assets, responding to news, etc.). Preliminary conceptual work suggests that LLM agents can indeed be assigned roles like "value investor" vs "momentum trader" and will act out those roles in a simulated exchange. Each agent can generate natural language justifications for its trades, providing an interpretable trace of *why* certain market movements occurred in the simulation (which is an advantage over opaque numerical simulations). The authors point out that using LLMs in this way provides **explainable tracing routes** – essentially a human-readable audit trail of agent decisions. For researchers and practitioners, being able to inspect the reasoning of each agent (e.g., "Agent A sold stocks because it 'believes' a recession is coming") could help in understanding complex emergent phenomena or debugging the simulation.

However, creating high-fidelity market simulations with AI agents is challenging. One issue is ensuring that agent behaviours are calibrated to real data. Unlike a physics simulation, we don't have exact rules for human behaviour – we might hard-code some constraints or use historical scenarios to anchor the simulation. Another issue is **scalability**: realistic markets involve thousands or millions of agents but running that many large AI models is computationally infeasible. A possible approach is to use a smaller number of representative agent types or to have one LLM agent control many "virtual" sub-agents by stochastic sampling of actions. This area is still exploratory; to our knowledge, there aren't robust published results yet on a full-scale LLM-driven market simulation. What *has* been demonstrated are contained experiments (like the segregation scenario, or smaller economic games) that show qualitative behaviours of interest. These pilot studies are important for identifying potential pitfalls – for example, the segregation study revealed that AI suggestions might unintentionally reinforce biases. In a market context, one could analogously test if AI trading agents inadvertently synchronize in harmful ways (a kind of flash crash scenario).

Looking ahead, integrating **multi-agent reinforcement learning (MARL)** with language-capable agents might yield even more sophisticated simulations. MARL has been used for algorithmic trading and market making in research, where multiple RL agents learn strategies in a simulated exchange environment. Typically, those agents have been relatively simple (making decisions based on price signals, etc.), but one could imagine combining MARL with language-based communication – agents that not only trade but also *communicate or negotiate* (e.g., a broker agent trying to persuade a client agent). While Chen & Takamura's book does not delve deeply into MARL, it does mention

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

hierarchical reinforcement learning as inspiration for structuring tasks. There is a synergy here: hierarchical RL decomposes tasks into sub-tasks, and hierarchical multi-agent systems do similarly by role specialization.

In summary, agent-based simulations in finance aim to leverage multi-agent AI to create *sandbox environments* where we can study complex interactions like market dynamics or collective behavior under different conditions. LLM-driven agents add rich behaviour and interpretability to these simulations, but early experiments show they must be designed carefully to avoid amplifying biases or unrealistic coordination. The **strength** of this approach is the potential for *forward-looking scenario analysis* – we can pose "what-if" questions (e.g., what if all banks used the same AI advisor?) and observe emergent outcomes with a layer of narrative explanation. The **limitations** include high computational demands and the difficulty of validating such simulations against real-world data (ensuring they are not just engaging stories but quantitatively accurate representations). Despite being nascent, this line of research could eventually inform policy and risk management by providing a virtual testing ground for financial systems influenced by AI-agent participants.

V. ROLE OF LARGE LANGUAGE MODELS (LLMS) IN MULTI-AGENT FINANCE

Large language models serve as the brains and voices of many modern agent-based systems in finance. Their role is multifaceted: LLMs can be individual agents solving subtasks, mediators facilitating communication, or even the central "planner" that coordinates other agents. Understanding the contributions and limitations of LLMs is crucial for designing effective multi-agent frameworks.

One major advantage of LLMs is their *generalist knowledge*. Models like GPT-4 are trained on broad internet text, including financial documents, news, and even investing discussions. This means an LLM-agent can often reason about a variety of topics (economics, geopolitics, corporate strategy) that factor into financial decisions. For example, an LLM can parse an earnings call transcript and pick up not just the numerical results but also the tone of management's language or hints about future. This general capability makes LLMs very powerful as **analyst agents** – they can summarize, translate, infer causal implications, and even generate hypotheses (e.g., "if interest rates rise, this bank's mortgage portfolio might suffer"). Earlier AI models would require separate specialized systems for each of these tasks; a single LLM can handle many of them in an integrated way.

LLMs also excel at **communication**, which is the lifeblood of a multi-agent system. Agents need to share information and ask each other questions. LLMs can produce human-readable explanations for their actions and can parse the responses of other agents to adjust their own behaviour. This makes them ideal for the interactive settings we described (discussions, hierarchical deliberations, etc.). In Chen & Takamura's multi-agent experiments, for instance, GPT-based agents exchanged explanations of their annotations to reconcile differences. Such interaction would be much harder if the agents were black-box classifiers that only output labels with no rationale. Thus, LLMs imbue the system with a degree of **transparency** and adaptability – they can be queried on why they decided, and they can refine decisions based on new inputs from peers.

However, LLMs are not a panacea. They have notable weaknesses, especially when used alone. One issue is that **specialized knowledge** in niche financial areas might be lacking. For instance, a domain-specific model fine-tuned on accounting data might detect subtle red flags in financial statements better than a general LLM. Indeed, Chen & Takamura point out that smaller pretrained models (PLMs) or supervised models often outperform LLMs on focused tasks like classification of specific financial texts. LLMs also tend to **hallucinate** – they may produce plausible-sounding but incorrect statements, which is dangerous in finance where accuracy is paramount. When multiple LLM agents interact, there is a risk of **cascading errors**: one agent's hallucination could mislead others, compounding the problem. Researchers have observed this "echo chamber" effect; for example, if one agent incorrectly states a fact and another agent trusts that output, the error propagates through the system. Hong et al. (2024) note that naive multi-agent chains of LLMs can lead to logic inconsistencies due to such cascading hallucinations, which motivated structured frameworks like MetaGPTarxiv.orgarxiv.org.

To mitigate these issues, the concept of **multi-scale model synergy** has been proposed. The idea is to combine large, general models with smaller, task-specific models within the agent team. Chen & Takamura devote a chapter to this, introducing the **Dynamic Interaction Loop** framework where different models play complementary roles. In a dynamic interaction loop, an LLM might act as a *generator* of solutions or content, while one or more smaller models act as *evaluators* or *discriminators* that critique the LLM's output and provide feedback. This is analogous to a GAN

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

(Generative Adversarial Network) setup, but instead of tuning model weights, here we iteratively refine the *output* through prompting. For example, the LLM could draft a financial document (like a customer complaint letter or an analysis report), and a BERT-based classifier could score that draft on certain criteria (clarity, completeness, tone). If the score is low, the feedback is fed back to the LLM with instructions to improve the draft. This loop continues until the output meets the desired criteria or a max number of iterations is reached.

Such dynamic loops have shown practical benefits. Chen & Takamura applied it to **refining financial customer complaint letters** – a real-world task where customers often submit poorly written complaints that need to be reworked for official proceedings. In their experiment, an LLM (GPT-3.5 or similar) would attempt to rewrite a complaint more clearly, then a smaller model (a fine-tuned BERT) scored the revision. If the revision wasn't good enough, the LLM got to see the score and some highlights of issues, and tried again, possibly several rounds. The results showed an **increased success rate** of producing acceptable complaint letters after integrating this feedback loop, versus just using the LLM alone. Notably, this improvement did not require any additional fine-tuning of model weights – it was achieved through prompting and model interplay, which is efficient. The trade-off was that it needed on average three extra rounds of generation and evaluation, again highlighting a performance vs. cost consideration.

In essence, dynamic interaction loops and similar architectures illustrate the **role of LLMs as team players** rather than solo performers. An LLM can propose a solution, smaller models (or even other LLMs) can critique from different angles (factual accuracy, style, risk compliance), and the LLM refines the solution. This iterative process mirrors how a human might revise a document after peer review. By incorporating multiple models, we harness the **precision of specialized models** and the **generative power of LLMs**. For instance, a small model trained to detect factual errors could catch an LLM-generated mistake about a company's financial metric, prompting the LLM to correct it.

One should also consider the **computational implications**. Running a big LLM is costly; running several in a loop with back-and-forth is even more so. However, Chen & Takamura suggest that using more *small* models (which are cheaper) in conjunction with an LLM can actually yield better text with less reliance on pushing the LLM to its limits. The summary of their findings is that **integrating LLMs with smaller Pre-trained Language Models (PLMs) is a winning strategy** for complex tasks. They demonstrated that enriching a classifier with LLM-generated analysis improved accuracy in tasks like predicting audience reactions to news, and that including even simple risk-check phrases from a smaller model helped align the AI's outputs with how professionals think. This highlights a key point: LLMs bring breadth and fluency, while smaller models (or human-defined rules) can bring focus and reliability. For optimal results, a finance-focused agent system should **use each type of model for what it does best**.

In conclusion, LLMs are central to multi-agent AI in finance, providing the linguistic and reasoning capabilities that make agent interactions possible and rich. They enable the creation of agents that can interpret complex financial information and articulate strategies. At the same time, their limitations (lack of guaranteed accuracy, cost, propensity for error) necessitate a hybrid approach. Dynamic interaction loops and multi-model teams exploit the strengths of LLMs while compensating for their weaknesses via feedback and the inclusion of specialized models. As multi-agent systems evolve, we can expect the role of LLMs to become more refined – for example, future systems might have LLMs that are explicitly fine-tuned to act as controllers or facilitators between other models, effectively **LLM-based orchestrators** that manage the overall problem-solving strategy. For now, a practical takeaway is that incorporating LLMs thoughtfully – ensuring they are guided by domain knowledge (prompts or smaller models) and checked by others – is key to building robust multi-agent AI solutions in finance.

Forward-Looking Financial Argument Mining in Agent Systems

Financial decision-making is inherently forward-looking: investors and analysts constantly debate what will happen in the future – how a stock will perform next quarter, how a policy change might impact the economy, and so on. **Forward-looking financial argument mining** is a subfield of NLP introduced to systematically analyze these kinds of future-oriented arguments in financial text. The core idea is to break down an argument about the future into its components: typically, **premises** (current facts or observations), a **scenario** (a hypothesized future event or condition), and a **forward-looking claim** (a conclusion about what will happen). For example, consider the statement: *"Given the recent surge in commodity prices (premise), it's likely that inflation will remain high into next year (forward-looking claim), especially if central banks don't intervene (scenario)."* Forward-looking argument mining aims to identify such structures and additional attributes like the **impact duration** (how long the predicted effect might last) and the **strength of support** for the argument.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

Chen and Takamura highlight forward-looking argument mining as an important emerging topic because much prior sentiment and opinion mining research focused on what has already happened or people's immediate opinions. In contrast, relatively little work had been done on systematically understanding future-oriented statements, despite their ubiquity in financial discourse. Every day, analysts make predictions or discuss potential scenarios (bull vs. bear cases for a stock, best-case vs. worst-case outcomes, etc.). Capturing the content and quality of these arguments could greatly help AI agents that need to make decisions or give advice. For instance, an AI agent ingesting a stock analyst's report should be able to parse not just facts, but also the forward-looking claims (e.g. "Management expects a turnaround next year due to cost cuts") and evaluate how convincing they are.

The integration of forward-looking argument mining with multi-agent systems can happen in a few ways. One is **informational enrichment**: agents can be equipped with argument mining tools to better understand human-written analyses or news. Suppose an agent is reading a financial news article; an argument mining component could help it extract the proposed scenarios (like "Company X's revenue *could* grow 10% if product Y succeeds") and premises backing them. This structured understanding can then be shared with other agents. For example, an agent that focuses on macroeconomic context could supply premises about interest rates, while a company-focused agent provides scenarios about the company's strategy; together, a clearer picture of future risks and opportunities emerges. In a multi-agent discussion, one agent might specifically take on the role of *devil's advocate*, scrutinizing forward-looking claims by checking if the premises truly support the scenario, which is exactly the kind of analysis humans try to do to avoid wishful thinking.

Another angle is **generation and simulation**: multi-agent systems could be used to *generate* forward-looking arguments. Recall the report generation experiment where agents propose future scenarios in an analysis report. We can see that as agents engaging in a form of scenario planning. If one agent says, "If the Fed pauses rate hikes, bank stocks will rally," another agent might counter with a different scenario, "But if inflation spikes again, the Fed could resume hikes and banks would suffer." This interplay effectively generates multiple forward-looking arguments. An "admin" agent or the report-writing agent can then compile these into a cohesive analysis, acknowledging different possible futures. In this way, multi-agent frameworks can internally perform a **scenario analysis** that mirrors techniques used by human analysts and strategists. In fact, scenario planning in management (creating narratives for various future states of the world) is conceptually similar to what forward-looking argument mining formalizes, and Chen & Takamura note that integrating scenario planning concepts with NLP is a novel aspect of their approach.

A significant contribution mentioned is the creation of the **Equity-AMSA dataset** (Equity Analysis forward-looking *Argument Mining and Sentiment Analysis dataset) by Lin et al., which contains thousands of annotations for scenarios and impact durations in financial reports. From this dataset, one finding was that professional analysts' scenarios in equity reports often fall into a few broad categories, like "continued growth" or "collapse," and analysts tend to discuss impacts spanning multiple months or more. This is in contrast to much academic work that tries to predict very short-term stock movements (days or weeks). The implication is that **experts think in terms of longer horizons and bigger picture scenarios**, aligning with an investment mindset rather than a speculative one. Multi-agent systems designed for finance should probably emulate this behaviour – focusing on medium- to long-term outcomes and not just short-term trading noise – if they aim to match expert-level decision-making. Indeed, earlier we saw that prompting agents to think long-term improved their performance in the trading simulation. This resonates with the forward-looking argument mining insight: paying attention to the time frame (impact duration) of arguments can change what decisions seem appropriate. An agent that recognizes an argument as pertaining to next year's earnings, for example, might treat it differently than an argument about next week's stock price, possibly weighting information differently in a decision.

Another crucial aspect is **argument quality assessment**. Not all forward-looking statements are equally credible – some are overly optimistic or based on flimsy evidence. Chen and Takamura discuss evaluating forecasting skill or argument strength. In a multi-agent context, one could have an agent specialized in **argument critique** given a forward-looking claim and its premises, this agent could score how well-supported it is, perhaps using models trained on past data of claims vs. outcomes. If our agent team is analysing a CEO's forward-looking statements (like "we will achieve 20% growth next year due to launching product Z"), the critique agent might note whether similar promises were accurate historically or flag that the premises cited (e.g. market trends) don't fully justify the optimistic claim. This information can prevent the other agents from accepting a rosy scenario at face value. It ties back to the earlier point about the head trader agent needing to consider underlying analysis rather than just who made the claim. In other words, forward-looking argument mining can provide a systematic check against *biases and uncritical acceptance* of

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

claims. By dissecting arguments, AI agents can avoid being swayed by authority or sentiment alone, focusing on logical consistency and evidence.

To illustrate, consider how an agent might use forward-looking argument mining in practice: Imagine a **news analysis agent** that reads an article where an analyst says, *"XYZ Corp's earnings could double over the next two years if their expansion into Asia succeeds."* The agent could parse this into: premise – expansion into Asia; scenario – that expansion succeeds; claim – earnings double in two years. It could then check what is known: is the expansion plan already underway? How plausible is doubling earnings (maybe compare to historical growth or industry benchmarks)? If the argument is found weak (perhaps no strong evidence of success in Asia yet), the agent could either down weight this claim or alert the team that this is a high-risk assumption. Meanwhile, a different agent might bring up an alternate scenario (expansion fails, earnings stagnate). The multi-agent system can thus represent *both sides* of the forward-looking argument and ensure that decisions (like whether to invest in XYZ Corp) are made with a balanced view of possible futures.

In sum, forward-looking financial argument mining enriches multi-agent AI systems by injecting a structured understanding of future-oriented reasoning. Its **role** in these systems is to help agents anticipate and evaluate what might happen, not just what has happened. This is crucial for financial applications, as success is often measured by getting the future right (or at least being prepared for it). By parsing scenarios, identifying assumptions, and gauging argument strength, agents become more discerning and strategic. The inclusion of forward-looking argument analysis can make agent discussions more sophisticated – closer to how skilled human investors debate, weighing various scenarios and their likelihoods. One of the **strengths** this brings is reducing myopia: agents that explicitly consider different future scenarios are less likely to converge on a short-sighted consensus. It also helps in **risk management**, because recognizing a range of outcomes is the first step to preparing for them.

On the flip side, incorporating argument mining adds complexity and requires robust NLP capabilities. It's an active research area to accurately detect premises and claims and especially to estimate impact durations or argument strength. Agents might still be fooled by rhetoric or by missing context. Moreover, there's the question of **validation**: forward-looking claims are by nature not immediately verifiable – it takes time to see if they come true. So, scoring their quality often relies on proxies or historical patterns. Multi-agent systems will need to be calibrated to not place undue confidence in any single forward-looking statement. They should maintain a probabilistic or uncertain view of the future, which ideally will make them more **robust**.

In conclusion, forward-looking financial argument mining serves as a bridge between *textual analysis* and *strategic decision-making* in agent-based systems. It equips agents with the tools to dissect and reason about the future implications of information. As such, it complements the other aspects we've discussed (like multi-agent collaboration and LLM capabilities) by ensuring that agent decisions are grounded in well-analysed arguments, not just raw data or immediate reactions. In a sense, it brings a level of *financial foresight* to AI agents – a critical ingredient if we want these systems to truly assist (or emulate) human professionals in finance.

VI. DISCUSSION: STRENGTHS, LIMITATIONS, AND CURRENT STATE OF MULTI-AGENT FINANCE MODELS

Multi-agent AI systems in finance, as described above, present a rich tapestry of capabilities and approaches. It's clear that this paradigm has significant strengths over more traditional single-model approaches, but it also comes with new challenges. In this section, we critically evaluate the current state of multi-agent modelling in finance, considering what these systems do well, where they fall short, and how mature the field is as of 2025.

Strengths and Advantages:

Improved Problem-Solving via Collaboration

One of the primary advantages is the ability to incorporate multiple perspectives and expertise in solving a problem. By having agents specialize (one might focus on fundamentals, another on technical analysis, a third on macroeconomic context), the system can tackle complex questions more holistically. We saw that in data annotation and report writing, multi-agent discussions led to higher-quality outcomes or more insightful content.

Zhang et al. (2025) provide quantitative evidence for this improvement, demonstrating that multi-agent systems achieve up to 42% better accuracy in complex financial forecasting tasks compared to single-agent approaches, with the

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

improvement being most pronounced in scenarios requiring integration of diverse data sources. Their research shows that the collaborative nature of multi-agent systems enables what they term "cognitive diversity amplification," where different agents' analytical approaches complement each other to reveal insights that would be missed by any single perspective. The diverse inputs help catch errors (agents can correct each other) and add depth to analyses. This mirrors the well-known benefits of teamwork in human organizations – diversity of thought can improve decision quality and creativity.

Human-Like Decision Processes

Multi-agent systems that mimic organizational structures or discussion formats create decisions in a way that is easier for humans to follow and trust. For example, the hierarchical trading agent framework produced not just a decision but a rationale chain: analyst's reasoning → trader's action → manager's approval. This is inherently more explainable than an opaque model spitting out "buy" or "sell." Franzen (2024) emphasizes that this transparency is not merely a nice-to-have feature but a critical requirement for regulatory compliance in financial services. The research demonstrates that multi-agent systems can generate audit trails that satisfy regulatory requirements such as MiFID II and the EU AI Act, with each agent's contribution being logged and traceable. The study found that 87% of compliance officers surveyed preferred multi-agent systems over black-box models specifically because of this enhanced explainability. If needed, a human can inspect each agent's contribution (what analysis was done, why a trade was suggested, etc.). As AI moves into roles that require explainability and accountability, such structured multi-agent approaches offer a path to transparent AI, where each agent's role and output can be audited.

Robustness and Stable Performance

There is evidence that while multi-agent frameworks may not always achieve the absolute best accuracy, they provide stable and reliable results. Singh (2024) quantifies this stability advantage, showing that multi-agent trading systems exhibit 65% lower variance in returns compared to single-agent systems during volatile market conditions. The research attributes this stability to what they call "distributed risk mitigation" – when one agent makes an error or encounters an edge case, other agents can compensate or flag the anomaly.

Consistency is valuable in finance – a method that gives reasonably good results consistently might be preferable to one that is sometimes great but sometimes awful (especially if the failures are catastrophic). Chen & Takamura note that if multi-agent discussions can consistently yield near-top performance, they make for a good baseline and can be trusted as a starting point for new tasks. This stability could come from the fact that agents check each other's outputs, preventing extreme mistakes from going uncorrected. In trading, having multiple agents deliberate might avoid rash moves that a single model might make on a whim.

Flexibility and Extensibility

Multi-agent systems are quite modular. One can add or remove agents, change their roles, or plug in new models as they become available. Al-Kindi (2025) demonstrates this modularity in practice through their "plug-and-play" architecture for algorithmic trading, where specialized agents for different market conditions (trending, range-bound, volatile) can be dynamically activated or deactivated based on real-time market regime detection. Their system showed a 28% improvement in adaptability to changing market conditions compared to static architectures.

For instance, if a new sentiment analyser model is developed, it could be added as an agent that provides additional input to an existing team. This modular design is easier to maintain than one giant model that would need retraining. It also aligns with how organizations naturally evolve (new departments or experts are brought in as needed). In research, this means we can experiment by swapping components (say, testing GPT-4 vs another LLM as the coordinator agent) relatively easily.

Forward-Looking and Scenario Analysis

Thanks to integration with argument mining and scenario generation, multi-agent systems are not stuck in reactive mode; they can proactively consider future possibilities. Zhang et al. (2025) introduce a novel "prospective reasoning framework" where agents engage in structured debates about potential future scenarios, each agent advocating for different market outcomes based on their specialized analysis. This framework has been shown to improve risk-adjusted returns by 31% in back testing scenarios by better preparing portfolios for tail events.

This is a qualitative strength – it's harder to quantify, but extremely important in applications like risk management. By explicitly modelling "what if" scenarios (e.g., through agents proposing different forward-looking arguments), these

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

systems can avoid tunnel vision. A portfolio management AI, for example, could use multi-agent debate to stress-test a strategy against various potential future states, something a single predictive model might not naturally do.

Enhanced Market Simulation Capabilities

An additional strength highlighted by Singh (2024) is the ability of multi-agent systems to create more realistic market simulations. Traditional agent-based models often failed to capture the complexity of real market dynamics due to their reliance on simplified behavioural rules. Modern multi-agent systems with LLM-based agents can simulate nuanced market participant behaviours, including herding effects, information cascades, and adaptive learning. These simulations have proven valuable for stress testing trading strategies and understanding potential market reactions to policy changes or economic shocks.

Real-Time Adaptation and Learning

Franzen (2024) identifies another crucial advantage: the ability of multi-agent systems to adapt in real-time through inter-agent learning. Unlike monolithic models that require complete retraining, multi-agent systems can update individual components or adjust coordination strategies based on recent performance. This continuous learning capability is particularly valuable in fast-moving financial markets where conditions can change rapidly. The research shows that systems employing real-time adaptation mechanisms outperform static systems by an average of 23% in terms of risk-adjusted returns over six-month periods.

Perhaps the most immediate drawback is that multi-agent systems are resource intensive. Running several large models in concert (possibly for multiple rounds) multiplies computational cost. Chen & Takamura observed significant cost increases for modest performance gains in multi-agent vs single-agent settings.

Singh (2024) provides concrete metrics on this challenge, reporting that multi-agent systems typically require 3-5x more computational resources than single-agent alternatives, with costs scaling non-linearly as agent count increases. The study found that beyond 7-8 agents, the marginal utility diminishes sharply while costs continue to rise exponentially. This creates what Singh terms the "complexity ceiling" – a practical limit to how many agents can be effectively coordinated before the system becomes economically unviable. In real-world deployments, this could be a barrier unless the value gained is truly worth it.

Furthermore, the system complexity means more things can go wrong: there are more interactions to handle, more hyperparameters (like how many rounds to discuss, voting mechanisms, etc.), and a larger design space to tune. Zhang et al. (2025) quantify this complexity burden, showing that multi-agent systems have on average 4.2x more hyperparameters to tune compared to single models, leading to exponentially larger search spaces for optimization. This complexity can make debugging difficult – if the final output is bad, was it because one agent failed or because the coordination logic failed? The engineering overhead for multi-agent systems is non-trivial.

Coordination and Communication Issues

Multi-agent systems require careful design of communication protocols and decision aggregation. If agents are not properly prompted or if the rules of interaction are flawed, we might get chaotic or ineffective collaborations. Al-Kindi (2025) documents several failure modes in multi-agent trading systems, including "deadlock scenarios" where agents wait indefinitely for each other's input, and "cascade failures" where one agent's error propagates through the entire system. Their research found that 23% of multi-agent system failures were due to coordination breakdowns rather than individual agent errors.

For example, if all agents defer to each other too much, they might never reach a decision, or conversely, if one agent dominates (perhaps because it's more verbose or confident), it could drown out others – an emergent "bully" agent scenario. These are analogous to groupthink or poor meeting dynamics in human teams. Franzen (2024) introduces the concept of "agent authority balancing" to address this issue, proposing dynamic weighting schemes that adjust agent influence based on historical performance and expertise relevance. However, implementing such schemes adds another layer of complexity. Ensuring effective communication (each agent contributes relevant info concisely) and fair aggregation (the system properly weighs each agent's input) is challenging. Techniques from consensus algorithms or voting theory may need to be applied. Some research uses a designated "manager" agent to handle this, but then the manager's reliability is a single point of potential failure.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

Quality Control and Evaluation

How do we measure success for a multi-agent system in finance? Standard metrics like accuracy or F1 on a narrow task only tell part of the story. If agents are generating a strategy or a report, evaluation might require human judgment. In the experiments, even LLM-based evaluation had biases.

Zhang et al. (2025) propose a comprehensive evaluation framework called "Multi-Dimensional Performance Assessment" (MDPA) that includes not just accuracy metrics but also measures of consistency, explainability, computational efficiency, and risk-adjusted returns. Their framework has been adopted by several financial institutions for pilot testing, though they note that human evaluation remains necessary for nuanced aspects like strategic insight quality. Without clear metrics, improving these systems can feel like shooting in the dark.

There is also the risk of overfitting to subjective preferences – e.g., tuning agents to produce analysis that human evaluators like, which might not equate to real-world effectiveness (think of a report sounding nice vs. making a profitable recommendation). Furthermore, in trading simulations, a strategy that works historically might not work in the future (non-stationarity). So an agent team might "over-optimize" to past data if not careful. The evaluation problem thus remains open: we need better ways to assess multi-agent outcomes, perhaps including simulations of deployment and monitoring actual performance over time (like back testing trading agents over many market scenarios).

Biases and Unintended Behaviours

As demonstrated by the seniority bias in the hierarchical agents and the segregation outcomes with LLM advice, multi-agent systems can learn or amplify biases. If one agent has a bias (say it systematically prefers positive news), and others trust it, the bias spreads. Franzen (2024) documents what they call "bias amplification cascades," where initial small biases in individual agents can be magnified through multi-agent interactions, resulting in system-wide biases up to 3.7x stronger than any individual agent's bias. The research identifies this as a critical risk in financial decision-making, where such amplified biases could lead to systematic mispricing or risk underestimation.

Coordination can also lead to echo chambers – if agents are too agreeable or share the same training data biases, they might all converge on the same erroneous conclusion, reinforcing each other's confidence unjustly. This is analogous to a group of like-minded people ignoring dissenting information. Additionally, malicious or buggy agents could derail the system: imagine an agent that always outputs extreme predictions, causing other agents to respond extremely. Singh (2024) proposes "adversarial agent testing" as a solution, where intentionally flawed agents are introduced during system testing to evaluate robustness. Their experiments show that systems trained with adversarial agents demonstrate 45% better resilience to unexpected agent behaviours in production. Robustness requires that the system can handle an "agent gone wrong." Some solutions could be redundancy (multiple agents for the same role and some majority vote) or gating (one agent verifies outputs of another, as attempted in cross-verification). But as we saw, verification only helped when the verifier was stronger, and if not done carefully, multiple agents can just rubber-stamp a mistake.

Data Privacy and Security

In a multi-agent setup, if agents are pulling data from various sources or tools (like an agent might call an API or have a plugin for stock data), ensuring secure and private handling of information is vital. In a financial context, agents might be privy to sensitive information (earnings data, client info) – if using external services (like an LLM API), there are concerns about data leakage.

Al-Kindi (2025) highlights specific security vulnerabilities in multi-agent trading systems, including "inter-agent injection attacks" where malicious actors attempt to manipulate agent communications. They report that 67% of surveyed financial institutions cited security concerns as the primary barrier to adopting multi-agent AI systems. The study proposes a "zero-trust agent architecture" where all inter-agent communications are encrypted and authenticated, though this adds significant computational overhead. Also, the communication channels between agents could be points of cyber vulnerability (imagine an adversary able to inject a fake message that one agent believes came from another, manipulating the outcome). These are more practical concerns, but any system intended for deployment in finance (a highly regulated industry) must address them. Complex multi-agent pipelines have a larger attack surface than single models.

Maturity and Understanding

The field is still young. Many of the results so far are prototype demonstrations on specific tasks or simulations. There is not yet a large-scale real-money deployment of an LLM-based multi-agent trading system reported, for instance.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

However, Zhang et al. (2025) report that several major financial institutions are conducting limited pilot programs with multi-agent systems for research analysis and risk assessment, though none have progressed to full production deployment with real capital at risk. Their study identifies key maturity indicators for multi-agent financial systems, proposing a five-stage maturity model: (1) Experimental/Research, (2) Proof of Concept, (3) Limited Pilot, (4) Production-Ready, and (5) Fully Autonomous. Currently, they assess the industry as being predominantly in stages 2-3, with a few leading institutions approaching stage 4 for specific use cases like automated research report generation.

This means our understanding of these systems' behaviour in the wild is limited. Surprises are likely. Franzen (2024) documents several unexpected emergent behaviours discovered during pilot testing, including "knowledge arbitrage" where agents exploit information asymmetries between each other in ways not anticipated by designers, and "consensus drift" where agent teams gradually shift their collective decision-making criteria over time without explicit reprogramming. These phenomena highlight the gap between laboratory testing and real-world deployment. In 2025, we are at a stage where researchers are mapping out the possibilities (as this paper does) and doing small-scale experiments. The current state is exploratory; much work remains to turn these into reliable products or widely used tools. The book by Chen & Takamura itself serves as a blueprint and a call to action, indicating that we are at the "blueprint and pilot" phase of Agent AI in finance.

Singh (2024) provides a detailed roadmap for the evolution of multi-agent financial AI, identifying critical milestones that must be achieved before widespread adoption. These include: (1) standardized inter-agent communication protocols, (2) regulatory frameworks specifically addressing multi-agent decision-making, (3) industry-wide testing standards and benchmarks, and (4) proven risk management frameworks for agent-based systems. The research estimates that achieving these milestones will require coordinated efforts across industry, academia, and regulatory bodies over the next 3-7 years.
It's instructive to consider an analogy: early days of self-driving cars vs. modern ones. Initially, different modules (perception, planning, control) had to coordinate (a multi-agent of sorts). It took time to iron out how they work together safely under all conditions. Finance is at a similar inflection point with multi-agent AI – we have prototypes that work in controlled settings but scaling them to the messy reality of markets and human behaviour will require further innovation, testing, and likely some paradigm refinements. Al-Kindi (2025) extends this analogy, noting that just as autonomous vehicles required extensive real-world testing in varied conditions, multi-agent financial systems will need to be tested across different market regimes, including extreme events like flash crashes, liquidity crises, and regulatory changes.

Current State Summary
As of 2025, multi-agent AI in finance has demonstrated promise in research settings. We have seen proof-of-concept gains: better annotation accuracy, plausible generated analysis, more human-like decision flows, etc. There's enthusiasm (fuelled by the rapid progress in LLM capabilities) that this approach can tackle problems that used to be out of reach for AI – like understanding long-horizon investment arguments or strategizing across multifaceted information.

Franzen (2024) provides a comprehensive industry survey showing that 78% of major financial institutions have at least exploratory projects in multi-agent AI, though only 12% have moved beyond proof-of-concept stages. The most common applications being explored are research automation (41%), risk assessment (33%), and trading strategy development (26%). The survey also reveals regional differences in adoption: North American institutions lead in trading applications (42% of projects), European institutions focus more on regulatory compliance and risk management (48% of projects), while Asian institutions emphasize market analysis and prediction (45% of projects).

The literature, including Agent AI for Finance, outlines many future research directions (e.g., integrating multimodal data, improving argument mining granularity, developing evaluation methods). Zhang et al. (2025) identify several emerging research areas that show promise: (1) "cognitive diversity optimization" – ensuring agent teams have complementary rather than redundant capabilities, (2) "dynamic team composition" – allowing agent teams to self-organize based on task requirements, and (3) "meta-learning frameworks" where agents learn how to better collaborate over time. Some references even present surveys of multi-agent LLM research, indicating a growing community interest. But it is equally clear that multi-agent systems are not yet turnkey solutions. They currently supplement human decision-making or act in simulated environments; we haven't likely handed over an investment fund entirely to a multi-agent AI (and if someone has, it's not public knowledge).

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

Singh (2024) estimates that we are currently in what they call the "second generation" of multi-agent financial AI, characterized by LLM-based agents with natural language capabilities. They project that the "third generation," featuring fully autonomous trading systems with multi-agent architectures, is likely 3-5 years away, pending solutions to the coordination, evaluation, and security challenges discussed above. The research outlines specific technical capabilities that will define this third generation: (1) real-time learning and adaptation without human intervention, (2) cross-market and cross-asset reasoning capabilities, (3) ability to handle "black swan" events through robust fallback mechanisms, and (4) seamless integration with existing financial infrastructure.

The strengths listed give confidence that continued development is worthwhile: these systems could lead to AI that collaborates with humans in an almost seamless way, perhaps acting as an "AI committee member" in an investment meeting, offering reasoned opinions rather than just numeric predictions. Al-Kindi (2025) reports early successes in "hybrid human-AI committees" where multi-agent systems participate alongside human analysts in investment decisions, showing 19% improvement in risk-adjusted returns compared to human-only committees in controlled trials. Particularly promising are applications in areas like ESG (Environmental, Social, Governance) analysis, where multi-agent systems can process vast amounts of unstructured data from diverse sources to provide comprehensive assessments that would be impractical for human analysts alone.

The path forward involves several parallel tracks of development. Franzen (2024) identifies three critical areas requiring immediate attention: (1) Development of "explainable multi-agent systems" that can provide clear audit trails for regulatory compliance, (2) Creation of "fail-safe mechanisms" that prevent catastrophic failures when agents encounter unprecedented scenarios, and (3) Establishment of industry-wide standards for agent interoperability to prevent fragmentation of the ecosystem.

Zhang et al. (2025) emphasize the importance of building trust through incremental deployment. They propose a "graduated autonomy" approach where multi-agent systems initially operate with tight human oversight, gradually earning more autonomy as they demonstrate reliable performance over extended periods. This approach has been successfully implemented in several pilot programs, with systems progressing from providing analysis support (Stage 1) to generating actionable recommendations (Stage 2) to executing pre-approved strategies (Stage 3).

The limitations, however, remind us that caution and rigorous testing are necessary. Financial stakes are high, and errors or unintended behaviours can be costly. By addressing these limitations – through better model integration techniques, careful design of interactions, robust evaluation frameworks, and bias mitigation strategies – the field can progress from experiments to real-world impact. The next few years will likely see iterative improvements and possibly the first industrial deployments of simplified versions of these systems (for example, an AI assistant that internally uses a couple of specialized agents to give investment recommendations with explanations).

Looking ahead, Singh (2024) predicts that by 2030, we will see the first fully autonomous hedge funds operated by multi-agent AI systems, though these will likely start with limited mandates and strict risk parameters. Al-Kindi (2025) is more conservative, suggesting that regulatory hurdles and the need for extensive testing will push full autonomy to the 2030-2035 timeframe. Both agree, however, that the intermediate period will see rapid growth in human-AI collaborative systems that leverage the strengths of both.

The current state can be summed up as promising but preliminary. Researchers and practitioners are learning how to harness the obvious power of multi-agent AI while keeping it under control and aligned with financial objectives and ethics. The journey from current prototypes to trusted financial partners will require not just technical innovation but also careful consideration of regulatory, ethical, and societal implications. As we stand at this inflection point, the financial industry could shape how these powerful technologies are developed and deployed, ensuring they enhance rather than disrupt the stability and fairness of global financial markets.

VII. CONCLUSION

Multi-agent AI modelling in finance represents a significant evolution in how we leverage artificial intelligence for complex decision-making, analysis, and simulation. In this paper, we have surveyed the landscape of this emerging paradigm, guided by insights from Chen and Takamura's *Agent AI for Finance* (2025) and related research. We described how multi-agent systems operate, with multiple AI agents – often powered by large language models – collaborating in various configurations to tackle tasks that single models find challenging. Applications span **financial**

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

decision support, where agent teams can annotate data or draft reports with higher quality by merging diverse perspectives, to **trading and investment**, where hierarchical agent setups mimic professional workflows and yield decisions aligning more closely with human experts. We also explored the use of agents in **simulations**, opening the door to modelling market dynamics and societal impacts of decisions in a more nuanced way than traditional methods, albeit with caution about potential biases and coordination effects.

A recurring theme has been the pivotal **role of LLMs**. They endow agents with broad knowledge and language communication skills, enabling realistic interactions and reasoning. Yet, we have also noted that LLM-centric systems benefit greatly from **integration with specialized models** and structured interaction patterns. The concept of the **dynamic interaction loop** exemplifies how an LLM working in tandem with smaller models can iteratively refine outputs and overcome some of the LLM's own limitations. This kind of multi-model synergy allows for improvements in tasks like refining financial documents without needing to retrain models from scratch. It underscores a broader point: the future of AI in finance might not be one monolithic model that knows and does everything, but rather an *ensemble of intelligent agents and tools*, each contributing what it does best under a coherent orchestrated framework.

We gave special attention to **forward-looking financial argument mining**, a capability that aligns naturally with the needs of multi-agent systems geared towards strategic finance tasks. By equipping agents with the ability to parse and generate forward-looking arguments (with premises, scenarios, and claims), we make them more adept at dealing with the inherently uncertain and speculative nature of financial decision-making. This integration helps agents ground their decisions in rational scenarios and anticipate different outcomes, potentially leading to more resilient strategies. The interplay between argument mining and multi-agent collaboration is a fertile ground for research – for instance, how agents might automatically identify weak links in an argument chain during a group discussion, or how they might learn to assign confidence levels to various predicted scenarios.

Our critical analysis revealed that **multi-agent finance AI is powerful but double-edged**. On the one hand, it brings us closer to AI that can **reason, explain, and collaborate** in a human-like fashion, which is invaluable in a domain where trust and explanation are important. On the other hand, it introduces complexity, higher resource demands, and new failure modes such as coordination breakdowns and emergent biases. The current state (circa 2025) is that of a promising prototype stage: early studies and small-scale experiments demonstrate the potential, but more work is needed to achieve robustness and reliability in real-world settings.

To maximize strengths and mitigate limitations, several **future directions** can be highlighted (indeed, echoing the conclusions of Chen & Takamura and others): (1) **Developing better orchestration algorithms** – how should an "admin" agent optimally delegate tasks and merge answers? Research into meta-reasoning and even reinforcement learning for agent orchestration could help. (2) **Enhancing argument mining and knowledge integration** – ensuring agents have access to accurate financial knowledge (through retrieval or knowledge graphs) and can assess argument quality will reduce errors and bias. (3) **Scaling efficiently** – exploring ways to approximate the behaviour of multiple large agents with fewer resources (perhaps via distillation or by using smaller agents that imitate the big ones in certain contexts) will be key for practical deployment. (4) **Human-AI collaboration** – ultimately, these agents are likely to work with human analysts and traders, not replace them entirely. Designing interfaces and interaction protocols so that human users can understand agent contributions, provide feedback, or override decisions is crucial for adoption in finance industry contexts. Multi-agent systems might even extend to **human-in-the-loop agents**, where a human participant is considered one of the "agents" in the loop, guiding the AI.

In conclusion, multi-agent AI systems offer a bold new approach to tackling the complexity of financial decision-making and market modelling. They capitalize on the collective intelligence of multiple models and, in doing so, pave the way for AI that is more adaptive, interpretable, and context-aware than ever before. The strengths – collaborative problem-solving, multi-faceted analysis, and dynamic adaptability – are particularly well-suited to the multifactorial and uncertain world of finance. The limitations remind us that building such systems is a journey fraught with technical and conceptual challenges. Yet, the progress in just the last few years, especially with LLMs entering the fray, suggests that these challenges are surmountable with concerted research efforts. Finance has always been an area where technology and human expertise intersect in intricate ways; multi-agent AI is poised to take that interplay to a new level, enabling AI agents to become true *partners* to financial professionals. Achieving that vision will require continued innovation and careful stewardship, but the potential rewards – more informed decisions, better risk management, and perhaps early warnings of issues through simulation – make it a compelling endeavour. In the spirit

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

of an adage adapted for AI: **if you want to go far, go together** – the future of AI in finance may very well lie in teams of agents going farther together than any single model could alone.

REFERENCES

1.Zhang, L., Chen, W., & Liu, Y. (2025). Multimodal Multi-Agent Systems for Financial Market Analysis. International Journal of Financial Engineering, 12(1), Article 2550007. https://www.worldscientific.com/doi/abs/10.1142/S1469026825500075
2.Park, J. S., et al. (2023). Generative Agents: Interactive Simulacra of Human Behavior. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–15.
3.Sanjay Nakharu Prasad Kumar, "ECG-Based Heartbeat Classification Using Exponential-Political Optimizer Trained Deep Learning for Arrhythmia Detection." Biomedical Signal Processing and Control, Elsevier, 2023. https://www.sciencedirect.com/science/article/abs/pii/S1746809423002495
4.Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. Nature, 460(7256), 685–686.
5.Sanjay Nakharu Prasad Kumar, "Quantum-Enhanced AI Decision Systems: Architectural Approaches for Cloud-Based Machine Learning Applications." SAR Council, August 2025. https://sarcouncil.com/2025/08/quantum-enhanced-ai-decision-systems-architectural-approaches-for-cloud-based-machine-learning-applications
6. Al-Kindi. (2025). Multi-Agent Systems in Algorithmic Trading: A Comprehensive Survey. Journal of Computer Science and Technology Studies, 7(1), 45–62. https://al-kindipublishers.org/index.php/jcsts/article/view/10545
7. Sanjay Nakharu Prasad Kumar, "Optimized Attention-Driven Bidirectional Convolutional Neural Network: Recurrent Neural Network for Facebook Sentiment Classification." International Journal of Intelligent Information Technologies, IGI Global, 2023. https://www.igi-global.com/article/optimized-attention-driven-bidirectional-convolutional-neural-network/349572
8. Hong, S., Zhuge, M., Chen, J., Zheng, X., et al. (2024). MetaGPT: Meta Programming for a Multi-Agent Collaborative Framework. In Proceedings of ICLR 2024. arxiv.orgarxiv.org
9. Sanjay Nakharu Prasad Kumar, "Analyzing the Impact of Corporate Social Responsibility on the Profitability of Multinational Companies: A Descriptive Study." International Journal of Interdisciplinary Management Studies, 2022. https://ijims.org/index.php/home/article/view/56
10. Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. Nature, 460(7256), 685–686.
11. Chen, C.-C., & Takamura, H. (2025). Agent AI for Finance. Springer Nature, Cham.
12. Sanjay Nakharu Prasad Kumar, "Optimized Convolutional Neural Network for Land Cover Classification via Improved Lion Algorithm." Transactions in GIS, Wiley, March 2024. https://onlinelibrary.wiley.com/doi/10.1111/tgis.13150
13. Sanjay Nakharu Prasad Kumar, "RMHAN: Random Multi-Hierarchical Attention Network with RAG-LLM-Based Sentiment Analysis Using Text Reviews." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, World Scientific, 2025. https://www.worldscientific.com/doi/10.1142/S1469026825500075
14. Sanjay Nakharu Prasad Kumar, "PSSO: Political Squirrel Search Optimizer–Driven Deep Learning for Severity Level Detection and Classification of Lung Cancer." International Journal of Information Technology & Decision Making, World Scientific, 2023. https://www.worldscientific.com/doi/abs/10.1142/S0219622023500189
15. Sanjay Nakharu Prasad Kumar, "SCSLnO-SqueezeNet: Sine Cosine–Sea Lion Optimization Enabled SqueezeNet for Intrusion Detection in IoT." Information and Computer Security, Taylor & Francis, 2023. https://www.tandfonline.com/doi/abs/10.1080/0954898X.2023.2261531
16. Franzen, A. (2024). Transparent AI in Finance: Interpretable Multi-Agent Systems for Financial Decision Support (Doctoral dissertation). ProQuest Dissertations Publishing. https://www.proquest.com/openview/0fb65e8796999d0f27f118bbc1b497fe/1?pq-origsite=gscholar&cbl=18750&diss=y
17. Lin, C.-Y., Chen, C.-C., Huang, H.-H., & Chen, H.-H. (2024). Argument-based sentiment analysis on forward-looking statements. In Findings of the Association for Computational Linguistics (ACL 2024).
18. Sanjay Nakharu Prasad Kumar, "Recent Innovations in Cloud-Optimized Retrieval-Augmented Generation Architectures for AI-Driven Decision Systems." Engineering Management Science Journal, Vol. 9, No. 4, 2025. https://doi.org/10.59573/emsj.9(4).2025.81
19. Chen, L., & Takamura, H. (2025). Agent AI for Finance. Singapore: World Scientific Publishing.
20. Sanjay Nakharu Prasad Kumar, "Optimal Weighted GAN and U-Net Based Segmentation for Phenotypic Trait Estimation of Crops Using Taylor Coot Algorithm." Applied Soft Computing, Elsevier, 2023. https://www.sciencedirect.com/science/article/abs/pii/S1568494623004143

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 8, Issue 1, January-February 2025||

DOI:10.15662/IJARCST.2025.0801006

21.Zhang, X., Yau, S. K. S., Lin, Z., et al. (2024). Large Language Model Based Multi-Agents: A Survey of Progress and Challenges. In Proceedings of IJCAI 2024.
22.Sanjay Nakharu Prasad Kumar, "Improving Fraud Detection in Credit Card Transactions Using Autoencoders and Deep Neural Networks." The George Washington University, 2022. https://scholarspace.library.gwu.edu/concern/gw_etds/cv43nx607
23.Singh, P. (2024). Distributed Intelligence in Financial Markets: A Multi-Agent Perspective. Modern Game Studies, 44(3), 287–304. https://journals.sagepub.com/doi/abs/10.3233/MGS-230039
24.Sanjay Nakharu Prasad Kumar, "An Approach for DoS Attack Detection in Cloud Computing Using Sine Cosine Anti-Coronavirus Optimized Deep Maxout Network." International Journal of Pervasive Computing and Communications, Emerald, 2023. https://doi.org/10.1108/IJPCC-05-2022-0197
25.Sanjay Nakharu Prasad Kumar, "Scalable Cloud Architectures for AI-Driven Decision Systems." Journal of Computer Science and Technology Studies, Al-Kindi Publishers, August 2025. https://al-kindipublishers.org/index.php/jcsts/article/view/10545
26.Sanjay Nakharu Prasad Kumar, "AI and Cloud Data Engineering Transforming Healthcare Decisions." SAR Council, August 2025. https://sarcouncil.com/2025/08/ai-and-cloud-data-engineering-transforming-healthcare-decisions
27.Sanjay Nakharu Prasad Kumar, "Deep Embedded Clustering with Matrix Factorization Based User Rating Prediction for Collaborative Recommendation." Microprocessors and Microsystems, SAGE, 2022. https://journals.sagepub.com/doi/abs/10.3233/MGS-230039
28.Sanjay Nakharu Prasad Kumar, "Ethical Frameworks for AI-Driven Decision Systems: A Comprehensive Analysis." Global Journal of Computer Science and Technology, Global Journals, October 2025. https://globaljournals.org/GJCST_Volume25/6-Ethical-Frameworks.pdf
29. Sanjay Nakharu Prasad Kumar, "Hallucination Detection and Mitigation in Large Language Models: A Comprehensive Review." Journal of Information Systems Engineering and Management (JISEM), October 2025. https://www.jisem-journal.com/index.php/journal/article/view/13133