



FACULTAD DE CIENCIAS ECONÓMICAS Y
EMPRESARIALES

Doctorado en Competitividad Empresarial y Territorial,
Innovación y Sostenibilidad

**Leveraging Sentiment Analysis and Artificial
Intelligence for Stock Market Prediction**

Autor
Alejandro Moreno

Dirigido por
Prof. Dr. Joaquín Ordieres-Meré

Madrid, mayo 2025



FACULTAD DE CIENCIAS ECONÓMICAS Y
EMPRESARIALES

Doctorado en Competitividad Empresarial y Territorial,
Innovación y Sostenibilidad

**Leveraging Sentiment Analysis and Artificial
Intelligence for Stock Market Prediction**

Author

Alejandro Moreno

Supervisor

Prof. Dr. Joaquín Ordieres-Meré

Madrid, May 2025

Resumen

Esta tesis explora la integración del análisis de sentimiento y las tecnologías de inteligencia artificial (IA) en la predicción del mercado de valores y la gestión de carteras. A través del análisis del sentimiento obtenido de diversas fuentes, como conferencias de resultados, informes de analistas, noticias y *tweets*, y empleando algoritmos de *Deep Learning* (DL) y modelos de lenguaje (LLMs), la investigación evalúa la efectividad de técnicas avanzadas de extracción de sentimiento y su impacto en la predicción de tendencias de precios de acciones.

El estudio también aborda la aplicación de estrategias impulsadas por IA para la gestión de carteras de acciones, comparando los asesores automatizados *Robo-advisors* con los nuevos agentes financieros autónomos (AFAs) basados en LLMs. Se pone un énfasis especial en el papel de la ingeniería de *prompts* para mejorar el rendimiento de los modelos de IA en la toma de decisiones financieras. A través de análisis empíricos y estudios de caso, la investigación demuestra el potencial del análisis de sentimiento y los enfoques basados en IA para mejorar las predicciones y optimizar el rendimiento de las carteras.

Los resultados contribuyen a un cuerpo creciente de literatura que conecta los campos de las finanzas y la IA, ofreciendo nuevas perspectivas sobre cómo estas tecnologías pueden moldear el futuro de la predicción del mercado de valores y la gestión de carteras. La tesis concluye con discusiones sobre las implicaciones éticas y las tendencias futuras en el uso de IA en finanzas, proporcionando recomendaciones para investigaciones futuras en este dominio en rápida evolución.

Keywords— Natural Language Processing, Large Language Models, Stock Market Prediction, Analyst Recommendations, Autonomous Agents.

Abstract

This thesis explores the integration of sentiment analysis and artificial intelligence (AI) technologies in stock market prediction and portfolio management. By analyzing sentiment derived from multiple sources, including earnings calls, analyst reports, news, and *tweets*, and using *Deep Learning* (DL) algorithms and language models (LLMs), the research evaluates the effectiveness of advanced sentiment extraction techniques and their role in predicting stock price trends.

The study also addresses the application of AI-driven strategies for stock portfolio management, comparing automated financial advisors (Robo-advisors) with new autonomous financial agents (AFAs) based on LLMs. Special emphasis is placed on the role of *prompt* engineering to enhance the performance of AI models in financial decision-making. Through empirical analysis and case studies, the research demonstrates the potential of sentiment analysis and AI-based approaches to improve predictions and optimize portfolio performance.

The results contribute to a growing body of literature that connects the fields of finance and AI, offering new perspectives on how these technologies can shape the future of stock market prediction and portfolio management. The thesis concludes with discussions on the ethical implications and future trends in the use of AI in finance, providing recommendations for future research in this rapidly evolving domain.

Keywords— Natural Language Processing, Large Language Models, Stock Market Prediction, Analyst Recommendations, Autonomous Agents.

Acknowledgements

This work, which now comes to an end, could not have been accomplished without the help of so many people who have accompanied me throughout these years.

First of all, I would like to thank my supervisor, Prof. Dr. Joaquín Ordieres-Meré for his professional generosity, one of the many things I have learned from him and hope to replicate in the future. Coordinator Prof. Dr. Carmen Valor has allowed me to grow while supporting me in everything I have needed.

I would also like to thank the entire Academic Committee of the Business and Territorial Competitiveness, Innovation, and Sustainability Program – CETIS, who have motivated me throughout the process. My thanks to Juan José Barbachano for his role as program manager and Dean Prof. Dr. Maria Teresa Corzo for her advice.

Next, I want to thank University of Geneva Professors Dr. Abraham Madroñal and Dr. Eduardo Solana for their support and generosity in hosting my exchange stay in Geneva.

To my family, whose support and patience have been the foundation of this journey.

And finally, to my children. I hope that the time this thesis has taken from them is compensated by the example of hard work, effort, and determination that I have tried to impart over the years.

Alejandro Moreno

Geneva, March 10, 2025

Credit Author Statement

Conceptualization, AM; methodology, AM; software, AM; validation, JOM; formal analysis, AM; investigation, AM; resources, AM; data curation, AM; original draft preparation, AM; review, JOM; visualization, AM; supervision, JOM; project administration, AM. All authors have read and agreed to the published version of the manuscript.

Use of GenAI

During the preparation of this dissertation, I used Elicit in order to help in the literature review. After using this Elicit, I reviewed and edited the content as needed and take full responsibility for the content of the publication.

Contents

1	Introduction	1
1.1	Background and Motivation	1
1.1.1	The AI revolution	2
1.1.2	AI in portfolio management	6
1.2	Research Questions	12
1.3	Contributions of the Thesis	13
1.4	Structure of the Thesis	15
2	Literature Review	17
2.1	AI in Stock Market Prediction	19
2.2	Sentiment Analysis in Finance	20
2.3	Autonomous Financial Agents	24
2.3.1	The rise of Large Language Models	24
2.3.2	Prompt engineering	24
2.3.3	Autonomous Financial Agents	27
2.4	Summary	31
3	Experiment design	33
3.1	Sentiment Analysis: IBEX Stocks Forecasting	36
3.1.1	Data	36
3.1.2	Methodology	41
3.2	Sentiment Jointly with Other Features: NASDAQ Stocks Forecasting . . .	44
3.2.1	Data	44
3.2.2	Methodology	47
3.3	AI Agent Forecast Based on Sentiment, Technical and Fundamental Anal- ysis: NASDAQ Stocks Forecasting	52
3.3.1	Fundamental and technical investment analyst	55
3.3.2	News investment analyst	55
4	Results	57
4.1	Sentiment Analysis: IBEX Stocks Forecasting	57
4.2	Sentiment Jointly with Other Features: NASDAQ Stocks Forecasting . . .	58
4.2.1	Portfolio performance evaluation	61

4.2.2	Portfolio evaluation	62
4.3	AI Agent Forecast Based on Sentiment, Technical and Fundamental Analysis: NASDAQ Stocks Forecasting	63
4.4	Summary	63
5	Discussion	67
5.1	Key Findings	68
5.1.1	Sentiment analysis: IBEX stocks forecasting	68
5.1.2	Sentiment, technical and fundamental analysis: NASDAQ stocks forecasting	70
5.1.3	Autonomous Financial Agents	71
5.2	Findings Implications	72
6	Conclusions and Future Research	75
6.1	Summary of Findings	75
6.2	Contributions to the Field	77
6.3	Limitations	78
6.4	Recommendations for Future Research	80
7	Practical implications	83
7.1	Stock Market Forecasting	83
7.1.1	Improved predictive performance based on sentiment	83
7.1.2	Role of LLMs and prompt engineering	84
7.1.3	Limitations and risks	84
7.2	Asset Management	85
7.2.1	Portfolio construction	85
7.2.2	Asset management efficiency	85
7.2.3	Portfolio risk management	86
7.3	Ethical and Regulatory Considerations	87
	Bibliography	89
	Appendix	109
	Literature review	i
A	Studies on AI in finance	i
	Agent results	iii
A	Results 16-27 Dec'24	iii
A.1	Prompting scenario 1	iii
A.2	Prompting scenario 2	iv
A.3	Prompting scenario 3	v
B	Results 23 Dec'24-3 Jan'25	vi
B.1	Prompting scenario 1	vi
B.2	Prompting scenario 2	vii

B.3	Prompting scenario 3	viii
-----	--------------------------------	------

List of Figures

1.1	Financial eras (Ghose, 2025)	1
1.2	AI evolution (Garlick et al., 2024)	3
1.3	ArXiv downloads as AI is an open collaboration project (source HBR-BCG)	4
1.4	AI CAPEX (Cook, 2024)	5
1.5	AI adoption by industry (Hatzius et al., 2024)	6
1.6	AI usage change forecast by industry (Hatzius et al., 2024)	6
1.7	Number of model downloads from Hugging Face (Kumar and Sharma, 2023)	6
1.8	Adoption rates are the highest within the largest (250+ employees) and smallest (1-4 employees) firms (Hatzius et al., 2024)	7
1.9	RAG solutions enable enterprises to deploy specialized use cases (Cook, 2024)	8
1.10	From Robo-advisors to AI Agents	9
1.11	Potential LLMs usages in portfolio management	11
1.12	AI in finance Bin Ren (2024)	13
1.13	AI in financial services (Ghose, 2025)	15
2.1	Prompt related papers evolution. www.promptingguide.com	25
2.2	Prompt engineering literature review term frequency. www.promptingguide.com	26
2.3	Multi-agent literature research gap	29
3.1	AI adoption according to geographies (Stanley et al., 2024)	34
3.2	Transformer architecture (Zhang et al., 2022b)	35
3.3	Selection of portfolio assets considering categories, variance and expected performance.	48
3.4	Discrete rebalancing strategy.	48
3.5	AI Agents vs LLM	52
3.6	Google searches on MAG-7 2022-2024	52
3.7	Multi-agent framework	53
4.1	LSTM	60
4.2	Transfor_1D	60
4.3	Transfor_4D	60
4.4	Performance comparison of different models using 10-fold cross-validation for GOOGLE class C assets	60

4.5	Prediction with different time windows for the asset value of Apple stock by using a multi-head Attention model with 16 heads and 8 layers	61
4.6	Variance vs λ	62
4.7	SR & DD factors	62
4.8	Alternative configurations for a portfolio. Each simulation corresponds to a different configuration of a portfolio. The simulation numbers correspond to the same experiment on both Figures 4.6 and 4.7	62
4.9	Anthropic Claude output based on price and news	64
4.10	Anthropic Claude price forecast based on price and news	64
4.11	10-day NVDA ARIMA, LSTM and AFAs forecast based on price technicals and fundamentals	65
6.1	Data is a new currency (Kumar and Sharma, 2023)	79
6.2	Processor capacity & related power consumption (Sheridan, 2025)	79
6.3	Model performance convergence: private & public (Stanley et al., 2024) . .	81
7.1	Exposure to AI automation by industry group (Hammond, 2025)	86
7.2	Quantitative performance is based in small gains (Fattouche et al., 2025) .	87

List of Tables

1.1	The Generative AI waves (Cook, 2025)	3
1.2	What kind of ML techniques do you find useful?	9
1.3	ML usage within your investment process?	9
1.4	What type of data you plan to apply to ML techniques?	10
1.5	New/alternative datasets evaluated last year	10
1.6	Generative AI uses cases for asset managers	10
2.1	Autonomous agents taxonomy papers 2021-23	27
3.1	European indexes size. October 2024 (Bloomberg)	37
3.2	Selected IBEX 35 companies: IBEX share, volatility (Beta) and industry breakdown	38
3.3	Sell-side equity reports for the selected IBEX companies issued by Barclays (BAR), Deutsche Bank (DB), JP Morgan (JPM), and Credit Suisse (CS)	38
3.4	Equity reports for the 10 selected IBEX 35 companies (2016–2022)	39
3.5	Top four most frequently used words in the analyst reports of the selected IBEX 35 companies, 2016–2022 (bag-of-words analysis)	39
3.6	Frequency analysis of the terms extracted from 3.5	40
3.7	Top 15 NASDAQ index components by market capitalization	45
4.1	Analyst recommendation and LLM extracted sentiment (%)	57
4.2	Analyst vs LLM extracted recommendation (Equation 3.7)	58
4.3	Stock price vs prediction (Equation. 3.4). 1% price range	59
4.4	Stock price vs prediction (Equation. 3.4). 3% price range	59
4.5	Stock price vs prediction (Equation. 3.4). 5% price range	59
4.6	Stock price vs prediction (Equation. 3.4). 7% price range	59
4.7	Comparison of stock price trend predictions by time horizon (AN: Analyst, FI: FinBERT, RO: Roberta, GP: GPT-4)	59
4.8	Simulation results	63
A.1	Literature review on stock-market forecasting using AI	ii
A.1	MSE Technical + Fundamental + News LSTM vs ARIMA	iii
A.2	MSE Technical + Fundamental + News AFAs vs ARIMA	iii
A.3	MAPE Technical + Fundamental + News LSTM vs ARIMA	iv

A.4	MAPE Technical + Fundamental + News AFAS vs ARIMA	iv
A.5	MSE Technical + Fundamental LSTM vs ARIMA	iv
A.6	MSE Technical + Fundamental AFAs vs ARIMA	iv
A.7	MAPE Technical + Fundamental LSTM vs ARIMA	iv
A.8	MAPE Technical + Fundamental AFAs vs ARIMA	iv
A.9	MSE News LSTM vs ARIMA	v
A.10	MSE News AFAs vs ARIMA	v
A.11	MAPE News LSTM vs ARIMA	v
A.12	MAPE News AFAs vs ARIMA	v
B.1	MSE Technical + Fundamental + News LSTM vs ARIMA	vi
B.2	MSE Technical + Fundamental + News AFAS vs ARIMA	vi
B.3	MAPE Technical + Fundamental + News LSTM vs ARIMA	vi
B.4	MAPE Technical + Fundamental + News AFAs vs ARIMA	vi
B.5	MSE Technical + Fundamental LSTM vs ARIMA	vii
B.6	MSE Technical + Fundamental AFAs vs ARIMA	vii
B.7	MAPE Technical + Fundamental LSTM vs ARIMA	vii
B.8	MAPE Technical + Fundamental AFAs vs ARIMA	vii
B.9	MSE News LSTM vs ARIMA	viii
B.10	MSE News AFAs vs ARIMA	viii
B.11	MAPE News LSTM vs ARIMA	viii
B.12	MAPE News AFAs vs ARIMA	viii

Abbreviations

<i>AFA</i>	Autonomous Financial Agent
<i>AI</i>	Artificial Intelligence
<i>AM</i>	Alejandro Moreno
<i>APE</i>	Automatic Prompt Engineer
<i>ASI</i>	Artificial Super Intelligence
<i>BERT</i>	Bidirectional Encoder Representations from Transformers
<i>BAR</i>	Barclays
<i>BofA</i>	Bank of America
<i>CAPM</i>	Capital Asset Pricing Model
<i>CAPEX</i>	Capital Expenditures
<i>CNN</i>	Convolutional Neural Networks
<i>CoK</i>	Chain-of-Knowledge prompting
<i>CoT</i>	Chain-of-Thoughts prompting
<i>CS</i>	Credit Suisse
<i>DB</i>	Database
<i>DB</i>	Deutsche Bank
<i>DD</i>	Drawdown
<i>DL</i>	Deep Learning
<i>EMH</i>	Efficient Market Hypothesis
<i>ESG</i>	Environmental, Social, and Governance
<i>FOMC</i>	Federal Reserve Open Market Committee
<i>GPT</i>	Generative Pre-trained Transformer
<i>GenAI</i>	Generative Artificial Intelligence
<i>GS</i>	Goldman Sachs
<i>HFT</i>	High-frequency trading
<i>IC</i>	Investment Committee
<i>ICADE</i>	Instituto Católico de Administración y Dirección de Empresas
<i>JPM</i>	JP Morgan bank
<i>JOM</i>	Joaquín Ordieres-Meré
<i>LLM</i>	Large Language Model
<i>LATAM</i>	Latin America
<i>LSTM</i>	Long-Short Term Memory
<i>MAG-7</i>	Apple, Microsoft, Nvidia, Tesla, Meta, Alphabet and Amazon NASDAQ stocks
<i>MACD</i>	Moving Average Convergence Divergence technical indicator
<i>ML</i>	Machine Learning

<i>NLP</i>	Natural Language Processing
<i>RAG</i>	Retrieval-Augmented Generation
<i>RSI</i>	Relative Strength Index technical indicator
<i>RNN</i>	Recurrent Neural Networks
<i>RoBERTa</i>	Robustly Optimized BERT Pretraining Approach
<i>ROI</i>	Return on Investments
<i>SR</i>	Sharpe Ratio
<i>SVM</i>	Support Vector Machines

Chapter 1

Introduction

1.1 Background and Motivation

The evolution of finance has been intrinsically linked to technological advancements for centuries (Figure 1.1). From the earliest forms of trade and banking systems to the modern landscape of algorithmic trading and artificial intelligence (AI), technological innovation has continuously shaped financial markets, institutions, and investment strategies. Each wave of technological progress has introduced new mechanisms for improving efficiency, reducing risk and expanding market accessibility.








	ERA	TRENDS	FINANCIAL FIRMS
	RENAISSANCE (1400s - 1500s)	Trade & Commerce in Europe	▪ Medici Bank (1397)
	EARLY MODERN (1600s - 1700s)	Maritime Trade, State Finance	▪ Bank of Amsterdam (1609) ▪ Bank of England (1694)
	INDUSTRIALIZATION (1700s - 1800s)	Industry, Infrastructure, International Trade	▪ Rothschild (late 1700s) ▪ Deutsche Bank (1870)
	COMPUTERS (1950s - 1980s)	Mainframe Computers, Proprietary Electric Networks	▪ Visa (1958) ▪ Mastercard (1966)
	INTERNET (1995 onwards)	Digital Economy, e-Commerce	▪ PayPal (1998) ▪ Alipay (2004)
	SMARTPHONE (2009 onwards)	Mobile Internet, App Economy, API Economy	▪ Nubank (2013) ▪ Revolut (2015)
	AI (2020s onwards)	Productivity, Personalization, Agentic AI Workflows	▪ AI Powered FinTechs ▪ Agile Incumbents

Figure 1.1: Financial eras (Ghose, 2025)

The origins of financial technology can be traced back to ancient civilizations, where rudimentary accounting and record-keeping methods facilitated commerce. The development of double-entry bookkeeping in the 15th century, pioneered by Luca Pacioli,

revolutionized financial management by providing a structured method for tracking assets, liabilities, and transactions. The introduction of the first stock exchanges, such as the Amsterdam Stock Exchange in 1602, further institutionalized financial markets, creating a structured environment for trading shares of the Dutch East India Company.

The Industrial Revolution of the 18th and 19th centuries marked a significant transformation in financial services, with the advent of telegraph-based communication allowing for faster information transmission across financial centers (J.P. Morgan, 2025). The introduction of ticker tape machines in the late 19th century enabled traders to receive real-time stock price updates, laying the foundation for modern electronic trading systems. The 20th century witnessed the emergence of computing technology, which significantly enhanced data processing capabilities, leading to the development of financial models, automated trading systems, and risk management tools.

The digital revolution of the late 20th and early 21st centuries has further accelerated the convergence of finance and technology. The proliferation of electronic trading platforms, high-frequency trading (HFT), blockchain technology, and AI-driven financial analytics has reshaped the landscape of financial markets. Large-scale data analysis, machine learning, and sentiment analysis have become essential tools for investors and financial institutions, allowing for more sophisticated decision-making and predictive modeling.

Artificial Intelligence (AI) and Natural Language Processing (NLP) are redefining how financial markets operate, with sentiment analysis and Large Language Models (LLMs) offering new insights into market trends. The emergence of autonomous financial agents (AFAs) represents the latest stage in this evolution, where AI-powered systems can analyze data, predict market movements, and execute trades without human intervention. As finance continues to embrace technological innovation, the boundaries between human expertise and machine intelligence are becoming increasingly blurred, raising both opportunities and challenges for investors, regulators, and financial institutions.

This thesis explores the latest developments in AI-driven financial analysis, particularly the role of sentiment analysis and LLMs in stock market prediction. By leveraging cutting-edge technologies, this research contributes to the ongoing dialogue on how finance and technology will continue to evolve in an era of unprecedented digital transformation.

1.1.1 The AI revolution

The pace of evolution of AI (Figure 1.2) brings us new and disruptive capabilities emerging on a continuous basis. AI technology is advancing so quickly that AI agents could soon handle end-to-end tasks regardless of industry: software coding, customer service, engineering design...and finance. BofA Global Research expects enterprise adoption of chatbots and copilots to accelerate over the next one to three years as capabilities improve and use cases expand (Table 1.1).

The evolution of Generative AI could transform into a revolution. While it may not lead to Artificial Super Intelligence (ASI) in the immediate future, we may be on the

Table 1.1: The Generative AI waves (Cook, 2025)

	1st wave	2nd wave	3rd wave
Technology	pre-GenAI	GenAI	Agentic AI
Time	1940-Nov'22	Nov'22-Oct'24	Oct'24 -
Application	NA	Chatbots & Copilots	Autonomous Agents

brink of a super cycle of innovation. The disruption AI brings in the near term should not be underestimated at this early stage. Although AI advancements might not result in ASI soon, the capabilities and use cases of AI agents—likely more comprehensive and disruptive than current chatbots and copilots—are expected to grow and evolve rapidly. AI’s progress will likely accelerate as computational power increases by 2-3 times annually without additional costs or energy consumption, with agent functionality expanding across computers, humanoids, industrial machines, lab instruments, restaurant equipment, and autonomous vehicles. In fact, we could be entering a super cycle of innovation, where increasingly powerful models generate more sophisticated applications, which, in turn, further enhance the models and capabilities of these apps.

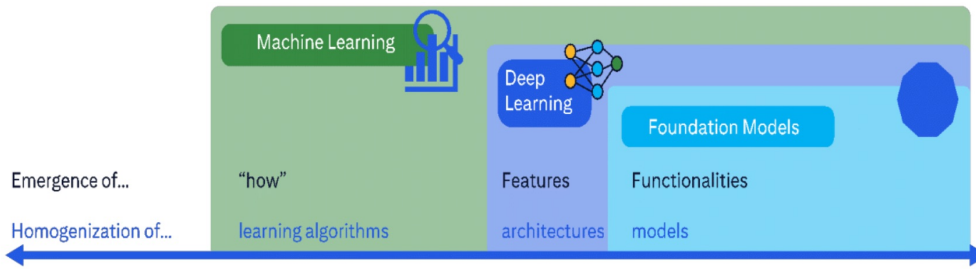


Figure 1.2: AI evolution (Garlick et al., 2024)

The capabilities of foundation models are advancing as they grow in size and benefit from new training techniques (Figure 1.3). Next-generation frontier models, which could cost over \$1 billion to develop (representing a 10x increase in training costs compared to the previous generation) are now reaching production. However, these larger and more powerful models also enable the creation of disruptive applications. As the underlying foundation models improve, chatbots for general tasks and copilots for specialized functions will likely become increasingly disruptive. These advances also pave the way for AI agents. While chatbots and copilots will eventually complement the development of agents, their functionality will continue to evolve and both consumer and enterprise adoption may accelerate as capabilities expand.

According to Cook (2025), GenAI usage among consumers and enterprises is expected to grow steadily in 2025, with a significant surge by 2026 as consumer products integrate GenAI features and pilot projects move into full production. AI tools are already capable of generating high-quality code, text, and images, producing research, analyzing trends,

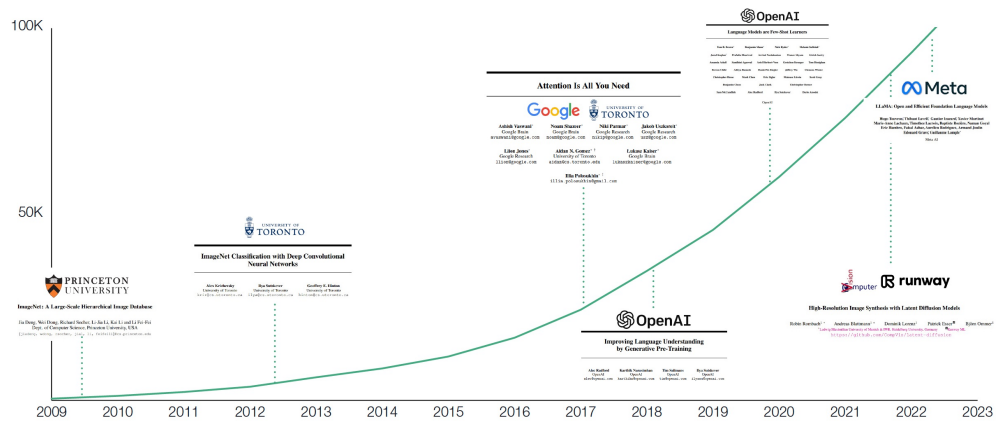


Figure 1.3: ArXiv downloads as AI is an open collaboration project (source HBR-BCG)

and optimizing processes. Yet, only 32% of US adults report using GenAI tools weekly and just 11% use them daily at work. While AI tools currently offer substantial efficiency and productivity gains, the next generation of these tools could bring even greater disruption. Adoption (Figure 1.5) is expected to accelerate in industries such as finance, transportation, and support services, especially as the cost savings become too significant to ignore (Figure 1.6). A report by Verdict (Williams, 2025) anticipates that GenAI will be the fastest growing segment within the AI market, potentially driving the next level of business process automation across various industries and reshaping existing business models. Similarly, Forrester's 2025 (Marrs, 2024) predictions highlight that marketing, sales, and product executives will face increased accountability as companies leverage GenAI to steer impactful initiatives. This suggests a strategic shift towards integrating GenAI to drive growth and efficiency. Additionally, the Federal Reserve's (Barr, 2025) noted that GenAI is poised to become a "general purpose technology" with widespread adoption leading to continuous improvements and productivity enhancements across various sectors. For the moment, much of the CAPEX has been deployed to infrastructure with only 23% of the \$21.8bn invested in private GenAI companies in 2023 flowed to companies developing applications (Figure 1.4).

These insights collectively indicate a trend where GenAI is expected to see steady growth in 2025, with a significant surge by 2026 as consumer products integrate GenAI features and pilot projects transition into full-scale production.

In 2018, approximately 60% of jobs were in occupations that didn't exist in 1940, highlighting the dynamic nature of the US economy. As foundation models and AI-powered applications advance (Figure 1.7), and as companies integrate AI tools into their products and processes, new jobs may emerge.

While much of the conversation around AI disruption focuses on knowledge-based jobs, such as those in writing and coding, where a study found a 21% drop in related freelance jobs globally within just eight months of ChatGPT's launch, less attention is

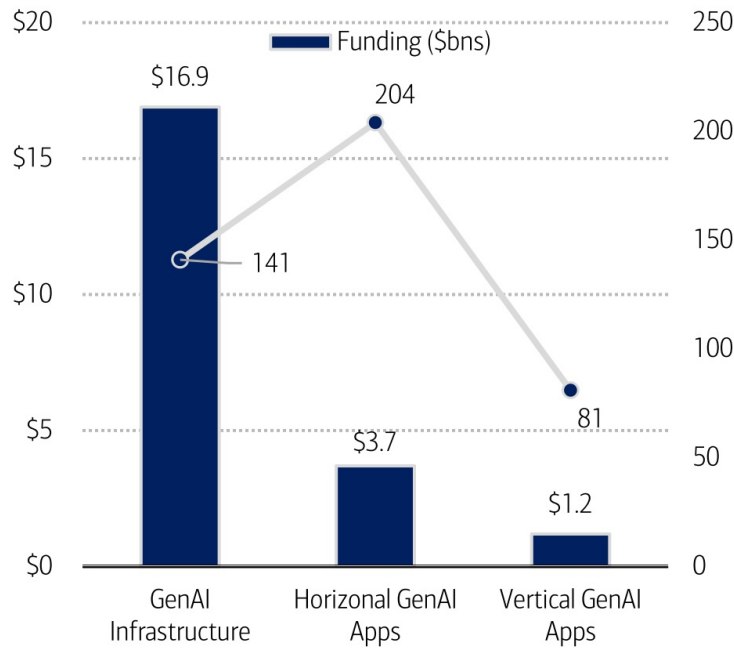


Figure 1.4: AI CAPEX (Cook, 2024)

given to how AI could impact physical labor occupations. Cook (2025) forecasts that the disruption driven by emerging robotic technologies could extend beyond customer-service agents, software engineers, marketing professionals, and medical workers to industries like warehousing, restaurants, construction, landscaping, and driving.

Robots have already been integrated into many industrial processes. For example, Amazon and DHL Supply Chain utilize 750,000 and 7,000 robots, respectively, with some robots capable of unloading boxes at twice the speed of humans and processing packages 25% faster.

Almost all the companies surveyed by Cook (2024) have developed an AI strategy, with only 4% not having one. The year 2024 marked the assessment of enterprise AI's return on investment (ROI), while 2025 is expected to be the year when enterprise AI is widely adopted. Enterprise GenAI applications are more advanced than simply entering random queries into ChatGPT, which would likely be too expensive for many enterprise applications that prioritize depth over breadth and may not result in significant efficiency or productivity improvements. This is due to the fact that the model behind ChatGPT is large and costly, exceeding the needs of many companies.

Enterprise GenAI applications will tap into small-language models (SLMs) and retrieval-augmented generation (RAG) solutions (Figure 1.9). They will enable enterprise-industry use case deployment with integrated internal, private, and live data.

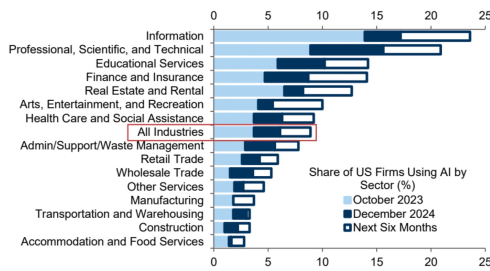


Figure 1.5: AI adoption by industry (Hatzius et al., 2024)

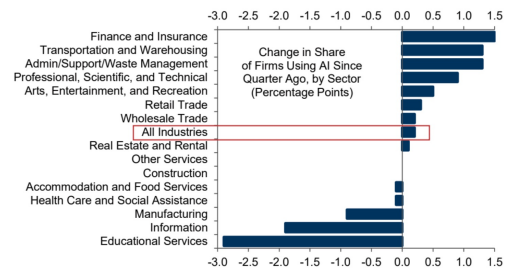


Figure 1.6: AI usage change forecast by industry (Hatzius et al., 2024)

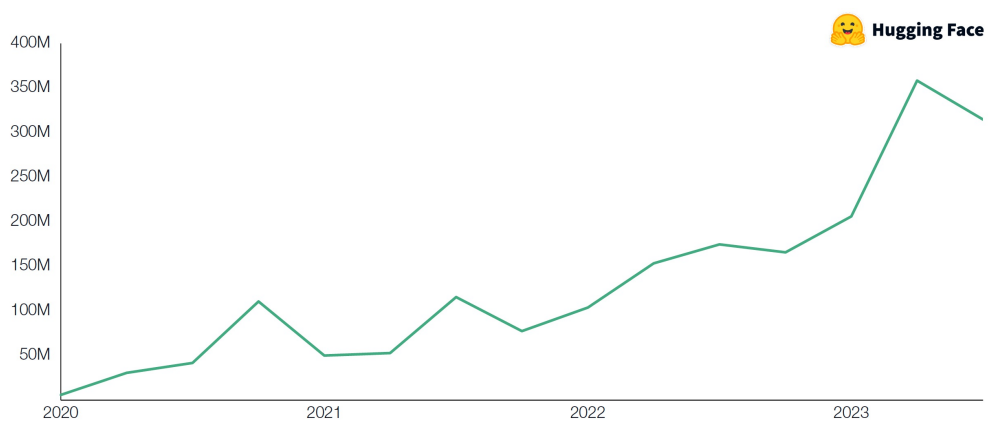


Figure 1.7: Number of model downloads from Hugging Face (Kumar and Sharma, 2023)

1.1.2 AI in portfolio management

Stock market fluctuations are driven by both publicly and privately shared information, including news articles, corporate disclosures, analyst reports, and discussions on social media platforms. By analyzing these diverse data sources, investors can potentially gain a competitive advantage. However, many existing studies still depend on traditional methods, which often struggle to effectively process and interpret unstructured data, such as textual information from these sources.

Portfolio management consists of making asset allocation decisions to create a portfolio with specific risk and return profiles. Artificial intelligence can help this process by improving fundamental analysis through quantitative or textual data analysis and developing innovative investment strategies Shöhnke M Bartram (2020). More sophisticated models better capture genuine nonlinear relationships, improving the performance of market timing strategies. This "virtue of complexity" (Jeff Cao, 2024) is demonstrated through three practical applications: timing the stock market, the bond market, and the long/short value factor. While these models offer real, though modest, performance

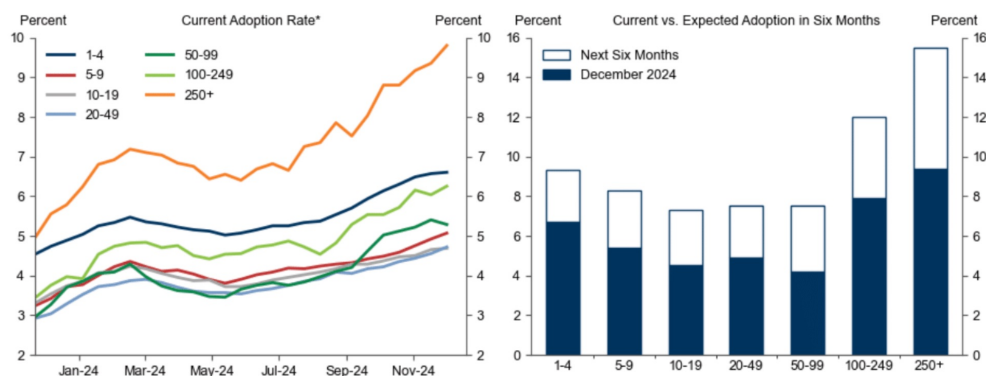


Figure 1.8: Adoption rates are the highest within the largest (250+ employees) and smallest (1-4 employees) firms (Hatzius et al., 2024)

enhancements, the results support the view that machine learning in return prediction leads to evolutionary, rather than revolutionary, gains in wealth.

A significant number of asset management firms are now using AI and statistical models to operate their trading and investment platforms Buchanan (2019). The relevant role of computers in trading was studied by Boehmer et al. (2021), Hendershott and Riordan (2013). AI techniques can address the limitations of traditional portfolio construction methods. AI can provide more precise asset return and risk estimates and address portfolio optimization problems with complex constraints (Mirete-Ferrer et al., 2022; Sen et al., 2021). The capability of AI to process large volumes of data enables the generation of trading signals. Algorithms can be trained to execute trades automatically based on these signals, contributing to the expansion of the algorithmic trading industry. In addition, AI techniques can reduce transaction costs by analyzing the market in real time and determining the optimal timing, size, and venue to execute trades. Technological advancements have dramatically changed the financial industry, along with many others, turning it into a space driven by digital innovation. This shift has had a significant impact on traditional portfolio management (Nguyen et al., 2023).

Robo-advisors are user-friendly, retail-focused systems that operate based on a set of pre-programmed instructions. According to Grealish and Kolm (2021), the future of financial advising will face challenges, shifting from rule-based investment management to goal-oriented investing and personalized portfolios. Research by Brenner and Meyll (2020) suggests that robo-advisors are becoming an appealing alternative for clients who are concerned about potential conflicts of interest or fraud with human advisors. Furthermore, a study by Tiberius et al. (2022) supports the growing dominance of robo-advisory services in the future, thanks to their affordability and ease of use.

The reliance on rule-based systems in robo-advisors, which follow models like modern portfolio theory or value-at-risk, can be a limitation in today's unpredictable investment environment. This approach makes portfolios more susceptible to extreme events, such as the COVID-19 crisis (Holtfort et al., 2022).

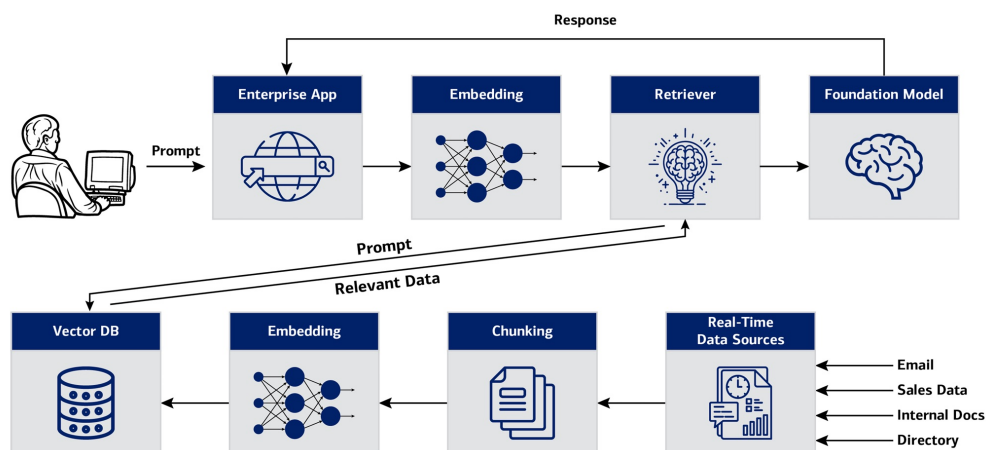


Figure 1.9: RAG solutions enable enterprises to deploy specialized use cases (Cook, 2024)

While robo-advisors have revolutionized portfolio management by offering cost-effective and automated solutions, their reliance on predefined rules limits their adaptability to evolving market conditions. This limitation has driven the creation of AI agents, which leverage advanced AI to not only analyze data but also engage with it autonomously, enhancing decision making in real time. In contrast to their rule-based predecessors, AI agents can learn from previous results and refine their strategies, positioning them as a powerful tool for the future of portfolio management.

The evolution of AI in portfolio management follows these stages (see Figure 1.10):

1. Robo-advisors: Rule-based. A static approach that relies on methods such as mean-variance optimization based on Markowitz's modern portfolio theory or risk-parity portfolios.
2. Generative Pre-trained Transformer (GPT) prompts: Task orientation. LLM-based portfolio management algorithms that use their transfer learning capabilities to interact with current situations (i.e., sentiment analysis of tweets or earnings transcripts). There is no feedback from the output to refine the prompt.
3. AI agents: Goal-oriented. Different agents interact among themselves in a feedback loop, retrieving (i.e., web scraping), analyzing (i.e., sentiment analysis) and sharing (i.e., impact of Federal Reserve minutes transcript sentiment with investment community *tweets*) to improve the outcome quality.

This framework breaks down technological progress into three main stages, each defined by major advancements in AI's ability to analyze data, interact with users, and learn from its environment to offer financial advice. AI agents not only process and analyze large volumes of financial data, but also enhance their decision-making abilities

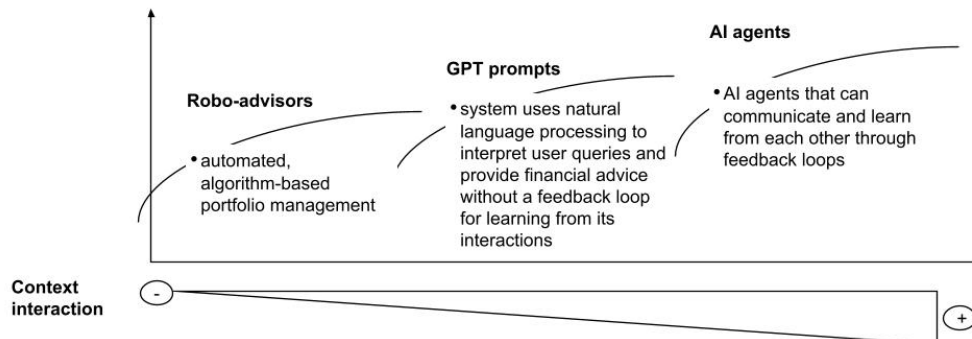


Figure 1.10: From Robo-advisors to AI Agents

over time by learning from the outcomes of their own actions as well as those of other agents.

Current AI usage in portfolio management is still limited as per industry conference held in Paris in October 2023 by the US bank JP Morgan (Chaudhry, 2023) (Tables 1.2, 1.3, 1.4, 1.5) surveying investments professionals. The use of AI algorithms in practice is merely a coding aid with the growing use of alternative datasets of unstructured data.

Skilled professionals would be even more productive with the help of AI and Generative AI as a recent study from Harvard University noted (Dell’Acqua et al., 2023). Those professionals not using Generative AI are less productive and potentially redundant. The paradigm is man and machine and not man vs. machine (Cao et al., 2021).

Table 1.2: What kind of ML techniques do you find useful?

Technique	%
Neural Networks	24
Trees/Random Forest	6
Gradient Boost	16
Linear models	9
Natural Language Processing	18
Generative AI / LLMs	25
Do not use	1

Table 1.3: ML usage within your investment process?

Usage	%
Enhance existing quant strategies	30
Build new quant strategies	35
Enhance portfolio construction	26
Make discretionary trades	5
Trade execution	4
Do not plan to use extensively	0

Table 1.4: What type of data you plan to apply to ML techniques?

Data	%
Fundamental data	13
Price/Market data	30
Text data, news, transcripts	38
Macro/economic data	5
Alternative data	12
Do not use	1

Table 1.5: New/alternative datasets evaluated last year

Frequency	%
None	25
1-3	40
4-10	17
More than 10	19

Despite their limitations, LLMs can be effectively applied to numerous aspects of portfolio management (Table 1.6) (Kaczmariski, 2023).

Table 1.6: Generative AI uses cases for asset managers

Research	Portfolio management
Synthesis of research and extraction of data from multiple sources	Scraping and transforming large amount of information from news for prediction of market movements or risk of securities
Drafting of market commentaries and outlooks	Generation of market or risk scenarios and synthetic historical and predicted data for backtesting
Drafting of investment committee documents	Proposal of new investment ideas and selecting relevant securities from a defined universe

The underlying principles of trading make LLMs a natural fit for enhancing the existing processes in these areas. Figure 1.11 shows the usual levers of investment decisions: fundamental analysis and technical analysis.

The context-based capabilities of LLMs and the vast training sets applied make it easier to identify undervalued stocks based on unstructured data or to detect price trends or even irregularities and anomalies.

Motivation for this study is to address the need to develop the understanding and capabilities of Generative AI in financial advisory and portfolio management. Due to the current traction of LLMs in different research fields but their relative scarcity in the financial domain, this study aims to setup the following research questions regarding AI agents in the financial domain (Autonomous Financial Agents or AFAs).

After 2008 crisis, risk management took a relevant part on the construction of portfolios. As financial assets and global markets grow more complex, traditional risk models may no longer be adequate for effective risk analysis. AI aids risk managers in validating and backtesting risk models. It can also extract information from diverse sources

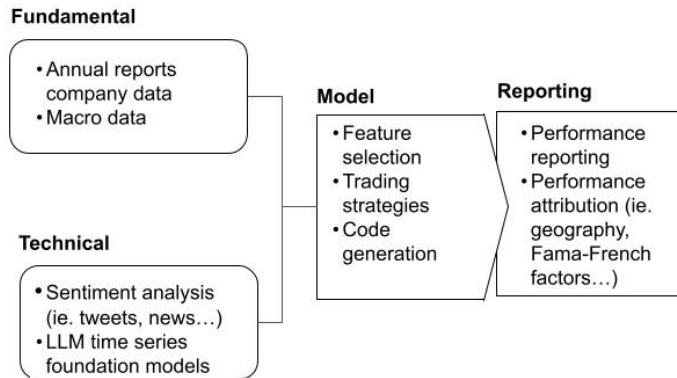


Figure 1.11: Potential LLMs usages in portfolio management

of structured or unstructured data, providing assessment of credit risk (Sirignano et al., 2018), market volatility (Zhang et al., 2022a), macroeconomic trends (Chen et al., 2023b), reputational risk (Myšková and Hájek, 2020) compared to traditional methods.

Sentiment reflects the collective opinions, emotions, or perspectives of a group. Investor sentiment serves as a method to gauge market sentiment. In their research, Dremman et al. (2001) emphasize how investor sentiment surveys have consistently revealed insights into investor attitudes over the years.

Research related to stock market prediction has attracted lots of attention due to the potential benefits of successful strategies. The unpredictability of the market, coupled with the vast amount of information and the large number of variables that could potentially affect stock values, as well as unanticipated noise, makes forecasting stock markets a difficult job (Henrique et al., 2019) for investors to make informed decisions (Bernales et al., 2021).

Price evolution is a confluence of buyers and sellers which economic decisions are driven by expectations. The Efficient Market Hypothesis (EMH) (Fama, 1970) states that available information is already reflected in the price and that investors are rational. The semi-strong form of EMH specifically suggests that all publicly available information is already incorporated into stock prices. Sell-side analyst research does not violate the semi-strong form of the EMH. Instead, it plays a role in the mechanism through which public information is incorporated into stock prices, supporting the notion of market efficiency.

The EMH has been challenged since its first release and different works by behav-

ioral economists and econometricians (Brown, 1999; Hsu et al., 2016; Shiller, 2003) have shown that there are reasons to question this hypothesis as shown by the development of consistently profitable factors based on market anomalies (Azevedo and Hoegner, 2022). Investor expectations may differ from rational forecast to build on personal beliefs subject to social influence. Although the EMH asserts that all relevant information is already incorporated into stock prices, analyst forecasts frequently challenge this premise.

Sentiment analysis involves applying natural language processing (NLP) and machine learning (ML) to evaluate public opinions or emotions related to financial markets. This analysis plays a key role in understanding market psychology and can be sourced from social media, news articles, earnings reports, or other text-based data. According to Guo et al. (2024), the impact of sentiment on price is more significant for small, non-profitable, low-tangibility, high-growth firms.

Previous studies showed that historical data jointly with textual data (Akita et al., 2016) or technical indicators with news sentiment (Li et al., 2020) improve the stock market forecast. Other studies highlighted the short-lived effect of new information on stock prices (Tetlock, 2007). Textual data focused on the firm under consideration do have more influence if the news are related to earnings and firm activity is close to daily activity (i.e., real estate, utilities) (Li et al., 2014).

This work investigates the effectiveness of LLMs and textual sentiment for stock evolution and portfolio management and is based on 3 studies:

- "Language Models Sentiment Extracted from Equity Analysts Reports as Stock Price Trend Predictor: Evidence from the IBEX Index". This paper examines how sentiment extracted from equity analysts reports using LLMs can act as a predictor of stock price trends, using the IBEX index as a case study ("*Economics Letters*", Elsevier). Accepted for publication.
- "Stock Portfolio Management Based on AI Technology". This paper investigates the effectiveness of AI-based models, including LLMs, in stock portfolio management, focusing on how these technologies can optimize investment strategies and improve risk management ("*Journal of Forecasting*", Wiley).
- "From Robo-advisors to Autonomous Financial Agents: Exploring the Role of Large Language Models and the Impact of Prompt Engineering". This study explores the evolution from rule-based robo-advisors to more sophisticated autonomous financial agents that utilize LLMs. It emphasizes the role of prompt engineering in enhancing the performance and decision-making capabilities of these models ("*Cogent Engineering*", Taylor & Francis). Accepted for publication.

1.2 Research Questions

Given the current popularity of LLMs in various research domains and their limited presence in the financial domain not to mention asset & portfolio management (Figure 1.12), this thesis seeks to formulate the following research questions:

- RQ1: Can sentiment extracted from equity analysts’ reports help to predict stock price trends? To what extent sentiment identified in earnings release or earnings calls by LLMs can improve its capability to explain short term stock price evolution?
- RQ2: Does Deep Learning time-series stock prediction with price, technical indicators and sentiment outperforms classical forecasting models?
- RQ3: How effective are autonomous financial agents in forecasting stock prices?



Figure 1.12: AI in finance Bin Ren (2024)

1.3 Contributions of the Thesis

This thesis highlights the significant contributions of combining forward-looking characteristics of sentiment analysis, price action, related technical indicators, and fundamental analysis in outperforming traditional forecasting models. While traditional forecasting methods often rely on historical data and fixed patterns, the forward-looking approach that incorporates sentiment provides a more dynamic understanding of market expectations and investor behavior. By analyzing sentiment data, one can anticipate market movements driven by psychological factors, such as investor optimism or fear, which traditional models might overlook. When this is combined with price action, which reflects real-time market conditions, and technical indicators that track historical trends and patterns, a more comprehensive and predictive framework is created.

Incorporating fundamental analysis, which examines the underlying economic factors affecting asset values, further enriches this integrated approach. This blend of sentiment, technical analysis, and fundamentals allows for more accurate and timely market predictions, as it incorporates both human emotions and objective financial data. Ultimately,

this combined approach can provide a superior edge over traditional models, which tend to rely solely on quantitative historical patterns, making them less adaptable to rapidly changing market conditions or unforeseen events. This research emphasizes the power of a multifaceted forecasting model that can better capture the complexities of modern financial markets.

This research aims not only to incorporate advanced forecast capabilities but advocating attributes such as the ability to regularly rebalance portfolio compositions. This approach seeks not only to address performance issues but also to accommodate dynamic changes in stock categorization across different categories.

The main contributions of this study are:

- **Limited Application of LLMs in the financial sector** despite the increasing popularity of LLMs in various disciplines, their adoption in the financial domain remains relatively scarce (Figure 1.13). This research aims to bridge this gap by exploring how language models can enhance stock market prediction.
- **Evolution from Robo-advisors to AFAs.** The study investigates the transition from traditional robo-advisors to AI-powered AFAs, which not only provide investment recommendations but also execute trades and manage risks.
- **Integration of sentiment analysis into predictive models.** Most traditional stock market prediction models rely on time-series data and quantitative indicators. This thesis introduces an approach that incorporates sentiment analysis extracted from analyst reports and financial news, adding a psychological layer of investor behavior that conventional models do not consider.
- **Differentiation of AI models for financial analysis.** The research explores how domain-specific financial models, such as FinBERT, compare with general-purpose models like GPT-4 in terms of accuracy and reliability for stock price prediction.
- **Combination of technical, fundamental, and sentiment-based factors.** The study argues that stock market prediction models can achieve superior performance by integrating technical analysis (price indicators), fundamental analysis (financial data) and sentiment analysis (NLP applied to news and reports). This holistic approach remains underexplored in the existing literature.
- **Impact of LLMs on market stability.** The thesis also examines how the increasing automation of financial decision-making through AI may influence market stability and investment dynamics.

In summary, this research fills a significant gap in the literature by integrating multiple advanced methodologies (NLP, LLMs, technical and fundamental analysis) into a framework that not only predicts stock market movements but also contextualizes them within investor psychology and the evolution of autonomous financial agents.








AGENTIC AI USE-CASES	WEALTH MANAGEMENT /RETAIL BANKING	CORPORATE BANKING	INSTITUTIONAL INVESTORS	INSURANCE
 PERSONALIZED OFFERS	<ul style="list-style-type: none"> Tailored financial advice adapting to client behavior/market trends Customized savings goals that adjust real-time to changes in income/expense 	<ul style="list-style-type: none"> Custom lending offers Optimized loan structures based on corporate profiles Dynamic pricing for corporate accounts 	<ul style="list-style-type: none"> Dynamic investment portfolios Bespoke investment plans 	<ul style="list-style-type: none"> Tailored insurance policies Dynamic loyalty offers based on profiles
 CUSTOMER ENGAGEMENT	<ul style="list-style-type: none"> Virtual financial assistants Tax planning agents Retirement guidance agents 	<ul style="list-style-type: none"> Financial planning agents Adaptive tax planning 	<ul style="list-style-type: none"> Custom research insights Real-time market alerts 	<ul style="list-style-type: none"> AI-driven customer support. Proactive policy updates, renewals
 OPERATIONAL EFFICIENCY	<ul style="list-style-type: none"> Automate routine tasks with context-aware workflows 	<ul style="list-style-type: none"> Streamline complex business operations Invoice processing and reconciliations 	<ul style="list-style-type: none"> Automated document review for fund performance 	<ul style="list-style-type: none"> Streamline claims with adaptive workflows
 RISK & UNDERWRITING	<ul style="list-style-type: none"> Dynamic risk profiling for credit approvals and investments Predictive models for default probability with real-time update 	<ul style="list-style-type: none"> Ongoing risk assessment and liability coverage recommendations with live data 	<ul style="list-style-type: none"> Portfolio diversification risk management Real-time hedging strategies 	<ul style="list-style-type: none"> Real-time underwriting models using health, climate, and behavioral data
 FINANCIAL FORECASTING	<ul style="list-style-type: none"> Insights for expenses, savings plans Predictive cashflows 	<ul style="list-style-type: none"> Live cashflow forecasts 	<ul style="list-style-type: none"> Real-time insights for timing of entry/exit in investments Investment performance insights 	<ul style="list-style-type: none"> Actuarial predictions for claims reserves Automated premium calculations and loss forecasting
 KYC/ ONBOARDING	<ul style="list-style-type: none"> Adaptive identity verification Real-time AML compliance 	<ul style="list-style-type: none"> Adaptive onboarding workflows Real-time sanctions monitoring 	<ul style="list-style-type: none"> Real-time due diligence Investor suitability analysis 	<ul style="list-style-type: none"> AI-assisted policy selection Adaptive identity verification
 FRAUD PREVENTION	<ul style="list-style-type: none"> Detect unauthorized activity using contextual data 	<ul style="list-style-type: none"> Monitor corporate fraud Anomaly detection in trade finance and payment invoices 	<ul style="list-style-type: none"> Insider trading detection Automated regulatory compliance 	<ul style="list-style-type: none"> Detecting claim anomalies Real-time fraud detection with contextual analysis

Figure 1.13: AI in financial services (Ghose, 2025)

1.4 Structure of the Thesis

The brief outline of this dissertation is as follows. A literature review of the related work is described in Chapter 2. Detail of the three experiments performed are described in Chapter 3. Results are included in Chapter 4 and we finally summarize results of this dissertation in Chapter 5 by presenting some of the assumptions we have considered and wrapping up the contributions of our work. Chapter 6 concludes and extends avenues for further research.

Chapter 2

Literature Review

The exploration of stock market prediction has garnered significant interest owing to the potential advantages offered by effective strategies. The complexity of the market, combined with the abundance of information and numerous variables that might impact stock values, alongside unforeseen disturbances, poses a challenging task in forecasting stock markets (Gu et al., 2023; Chopra and Sharma, 2021; Henrique et al., 2019) and for investors to make informed decisions (Bernales et al., 2021).

The Efficient Market Hypothesis (Fama, 1970)) and the Random Walk theory (Malkiel, 2003) state that market returns cannot be precisely predicted. This belief is based on the presumption that investors are rational and that everyone has access to all data from the public market. The EMH has been frequently challenged and different works by behavioral economists and econometricians. Hsu et al. (2016); Brown (1999) have shown that there is reason to question this hypothesis, as shown by the development of consistently profitable factors based on market anomalies (Azevedo et al., 2024; Azevedo and Hoegner, 2022). Price evolution is a confluence of buyers and sellers whose economic decisions are driven by expectations. These expectations are not only rational, they are also built on personal beliefs, often faced with social influence. Since the 1990s, empirical data published by behavioral finance specialists have shown that investor psychology is what drives the stock market (Daniel et al., 1998).

The Capital Asset Pricing Model (CAPM) has been subject to extensive scrutiny since its inception in the 1960s (Neslihanoglu et al., 2017). Specifically, the effectiveness of the market capitalization weighted index has been challenged, leading academics and professionals to propose various alternative investment options.

Traditional models for predicting market behavior were based on fundamental analysis (company trends) or technical analysis (price trends) (McMillan, 2016). The continuous development of new models and concepts in this area has contributed significantly to the evolution of modern financial theories. AI models, which are now widely used as analytical tools, have recently made notable advances in portfolio optimization, becoming increasingly sophisticated (Lin and Taamouti, 2024; Karathanasopoulos et al., 2017; Guerard et al., 2015).

Portfolio management encompasses the tasks of choosing stocks to incorporate into a

portfolio, determining the allocation of capital to each stock, and identifying the moments that require portfolio rebalancing. Throughout this process, it is crucial to take into account the investor's risk tolerance, as certain investors are more open to assuming higher levels of risk in hopes of achieving potentially higher returns. In recent years, several groundbreaking contributions have emerged in the field of stock portfolio optimization, incorporating advanced machine learning techniques and innovative algorithms. These contributions have significantly improved the precision and efficacy of stock market predictions and portfolio management.

Academics work to balance the seemingly conflicting goals of maximizing returns and minimizing risks to optimize diversified portfolios. They create new models and estimation techniques to improve the Markowitz model. The use of network theory in portfolio management provides a novel approach to address asset allocation challenges. In this framework, assets are viewed as nodes linked by their relationships (Tumminello et al., 2005). One notable advancement is the mean Value-at-Risk (VaR) model combined with AdaBoost prediction, which has demonstrated superior performance in the optimization of an international stock market portfolio compared to other machine learning regression models (Behera et al., 2023). This model improves the predictability and reliability of financial returns, providing investors with robust tools to manage portfolios. Another significant development is the use of long- and short-term memory (LSTM) Deep Learning models to predict future stock prices with high precision. According to Jang and Seong (2023), by integrating modern portfolio theory with deep reinforcement learning, it is possible to outperform traditional algorithms in terms of Sharpe ratio, annualized return and minimum drawdowns (Dacorogna et al., 1999). This combination exploits the strengths of both theory and practice to achieve optimal portfolio performance. However, portfolio optimization currently lacks the flexibility to create a parametric selection of stocks from diverse categories beyond their sectors, hindering the ability to assess performance under constraints related to diversification, adherence to ESG principles and other criteria.

In the context of optimal portfolio construction, research in the timber and forest sector demonstrates that incorporating Environmental, Social, and Governance (ESG) scores into investment strategies can achieve risk-adjusted returns comparable to traditional investments. This supports the feasibility of socially responsible investments in maintaining financial performance while promoting sustainability (Arreola Hernandez and Al Janabi, 2020; Löff et al., 2023). Furthermore, studies in Latin America confirm that portfolios with high-ESG stocks outperform those with low-ESG or non-reporting companies, showcasing the benefits of ESG criteria in investment decision making (Useche et al., 2024).

Study researched literature with key terms related to stock market, analysts, sentiment, AI, AFAs, DL, LLM, ML, NLP and the likes for the period 2015-2025 for a total of roughly 1000 documents classified and referenced in *Mendeley* database under the following three sections:

2.1 AI in Stock Market Prediction

In recent years, AI and Machine Learning have been used to analyze market trends and help investors make better decisions. Finance has been close to AI since the early stages Bickley et al. (2022).

Machine Learning models and more precisely Deep Learning (DL) models (DL is a neural network with two or more hidden layers) like LSTM have the advantage of being able to handle non-linearity. Non-linearity of financial markets is a basis for the application of ML techniques. To prevent the excessive proliferation of features, machine learning has introduced techniques such as *feature engineering* that can account for time evolution while maintaining an inherently nonlinear behavior (i.e., LSTM) (Bao et al., 2017; Chaweewanchon and Chaysiri, 2022; Dingli and Fournier, 2017; Sezer et al., 2020). One of the advantages of machine learning is its ability to learn from examples through Supervised Machine Learning eliminating the need to manually select input variables.

Vaswani et al. (2017) introduced in their seminal paper the Attention mechanism into LSTM, applied to the NLP domain. On the other hand, one of the first studies to suggest the use of an LSTM + Attention mechanism for multivariate prediction of time series is Shih et al. (2019). Li et al. 2018 used a multi-input Attention LSTM to separate useful information from negatively associated elements and eliminate their noise. Qiu et al. (2020) used an Attention mechanism to denoise historical stock data. Similarly, Li et al. (2022) integrated a Transformer encoder with an Attention mechanism and social media *tweets* to predict stock movements, utilizing social media only to identify the most relevant stocks for analysis. Heaton et al. (2017) showed that the incorporation of Attention layers into LSTM enhances prediction capabilities compared to previous Deep Learning models. This approach has been effectively applied to different sectors of the Indian stock market, offering valuable predictions that support the optimization of stock portfolios (Sen et al., 2021).

The studies cited in Table A.1 and those already stated, suggest that combining textual information (news items) with pricing (historical prices and technical indicators) improves price prediction compared to a price-only or text-only strategy (Selvin et al., 2017). Machine Learning models such as SVM, Random Forest, and Decision Trees have demonstrated high accuracy (90%) in predicting stock trends when combining sentiment data with traditional market indicators (Patil et al., 2022). However, Deep Learning approaches, particularly LSTM, deliver the best performance, achieving an accuracy of 92%.

Many funds now use algorithms to analyze transcripts of earnings calls and company presentations, extracting signals from executives' language through NLP (Kirtac and Germano, 2024). The automated retrieval of quarterly and annual reports in the United States conducted through an algorithmic scraping process rather than human reading, surged from approximately 360,000 in 2003 to 165 million in 2016 (Cao et al., 2020). Authors noted that companies are eager to present their business in the most favorable manner possible. Consequently, they have progressively enhanced the machine readability of their reports, such as adjusting table formatting in response to the dominance of

machine analysis. Those NLP algorithms process the information and trigger buying or selling orders without human intervention in fractions of seconds.

The technology available to hedge funds for word scraping is increasingly available for corporates that avoid using words with negative sentiment as qualified under the finance focus dictionary (Loughran and McDonald, 2020) or the broader Harvard Psychosociological dictionary. That has led to standardized written reports AI-proven and the dominance of positive messages that do not discriminate on the prospects of the company. Cao et al. (2020) discovered that since 2011, companies have subtly adjusted the language used in reports and the way executives communicate during conference calls. This is done to avoid using words that could raise concerns when detected by machine listening systems. Dictionary-based methods (Frankel et al., 2022) to extract sentiment have been overtaken by Transformer-based Large Language Models. There is a cat-and-mouse race between companies and buy-side investors. The latter can promptly modify the model to notice any deliberate effort to convey positivity or negativity in response to news.

2.2 Sentiment Analysis in Finance

The crowd of traders' mood is called market sentiment. Defining investor sentiment is a long standing question (El Alaoui et al., 2020; Mishev et al., 2020; Zhang et al., 2018b). Since the middle of the 20th century when Schumpeter and Keynes (1936) first suggested that markets are influenced by "animal spirits" of investors which cause prices to diverge from fundamentals, investor sentiment has been a widely accepted concept. This concept was later expanded by Long et al. (1990), who demonstrated how changes in sentiment could cause noisy trading and volatility spikes. The goal is to assess whether the sentiment will likely drive stock prices up or down through any of the following sentiment analysis methodologies:

- Lexicon-based sentiment analysis: This technique uses predefined lists of words with assigned sentiment scores (positive, negative, or neutral). The sentiment of a text is determined by summing the scores of the words it contains.
- Machine Learning-based sentiment analysis: In this method, machine learning models are trained on labeled datasets to recognize sentiment in new text. Common techniques include Support Vector Machines (SVM), Naive Bayes classifiers, and logistic regression.
- Deep Learning-based sentiment analysis: Advanced approaches leverage deep learning algorithms, such as Recurrent Neural Networks (RNNs) and LSTM networks, to capture intricate relationships and dependencies within the text, thereby enhancing sentiment detection accuracy.

People learn from others' behavior not just through price movements, but also by following the news, reading articles, and engaging in discussions on social media platforms like Twitter, Reddit, and others (Yang et al., 2022). We often reevaluate our beliefs by

testing them within a social context. Behavioral finance literature typically views the *herd* as homogeneous, giving less focus to individuals with significant influence (Fallahgoul and Lin, 2021).

Ignoring the potential effect of sentiment and simply following price action, the forecasting problem may be just that of time series forecasting (Adebiyi et al., 2014). As per Mohan et al. (2019), predicting stock prices based on either historical data or textual information alone has proven to be insufficient. The difficulty on stock market forecasting comes from the non-linearity of the stock price evolution.

There is still a debate on whether sentiment influences returns or vice versa, as discussed by several authors (Jiang, 2021; Shen et al., 2023; Wang et al., 2006). If sentiment influences the market, optimism within the investment community could lead to a bull market, while pessimism might trigger a bear market. Recognizing the current market sentiment can aid in predicting future market trends and create opportunities for sentiment-based trading. Saetia and Yokrattanasak (2022) argued that sentiment was not significant in stock selection or portfolio construction, but in their study, sentiment was represented by the presence of specific keywords, which may not fully capture the true sentiment. Barberis et al. (1998) study indicates that investor sentiment is inconsistently linked to stock prices. It revealed that investor sentiment tends to react underly to company information, such as earnings announcements, while overreacting to a sustained history of extreme performance, whether positive or negative. The author (Baker and Wurgler, 2006, 2007) researched the sentiments on individual firms and the market as a whole. Thorp (2004) mentions the lagging characteristic of sentiment and irrational emotions that drive prices. The study of Lemmon and Portniaguina (2006) found relationship between sentiment and returns for small stocks. (Kaplanski et al., 2015) found a link between positive sentiment, inclination to buy stock and subsequent returns. Study from Ling et al. (2014) found that private markets are susceptible to prolonged periods of sentiment-induced mispricing. Chatterjee and Perrizo (2016) discovered that when sentiment is low, the future returns of growth stocks outperform those of value stocks. Conversely, when sentiment is high, the future returns of growth stocks are lower than those of value stocks.

Sentiment analysis based on NLP is part of the ML effort and has the potential to enable human-like language interpretation in various applications, including the analysis of stock-related news and earnings reports. Analyzing language is analyzing sentiment. Sentiment analysis is an NLP technique that identifies the polarity of a given text such as positive, negative and neutral. The business potential of conversational AI technologies in finance has yet to be discovered (Yue et al., 2023).

The literature on NLP has found that Transformer models achieved remarkable performance in language modeling, surpassing previous dictionary-based algorithms. The release of Large Language Models like Bidirectional Encoder Representations from Transformers (BERT) Devlin et al. (2018) and Generative Pre-trained Transformers (GPT) Radford et al. (2019) was a leap ahead in NLP. Sentiment extraction through LLM is far more insightful than previous sentiment analysis techniques such as bag-of-words. The already mentioned seminal paper from Vaswani et al. (2017) on the Attention mechanism

launched a race on size of transfer learning models based on Transformer architectures. To mention a few: GPT-3 Brown et al. (2020), Gopher Rae et al. (2022), Bloom Lounen et al. (2022), GPT-NeoX-20B Black et al. (2022) and META OPT-66B language model Zhang et al. (2022c).

Reviewing existing literature on sentiment analysis in finance, most of papers employ lexicon-based approaches that concentrate on extensive word corpora. The authors Wisniewski and Yekini (2015) reviewed 1'262 annual reports (2006-2012) of 209 listed UK firms to extract sentiment using a lexicon-based approach. They counted word frequency under these three categories: praise, concreteness and activity to forecast future stock performance. They found weak correlation coefficients with stock price evolution (0.0708 to 0.08027). Authors Loughgran et al. (2011) studied 50'155 annual reports 1994-2008 for US based companies and found that lexicon-based sentiment classification may not extract sentiment properly when applied to the financial domain therefore reducing their meaningfulness. Previously, author Li (2006) studied 10'000 US annual reports filings through word frequency counting and found that reports with negative sentiment were followed by share price negative performance in a 12-month horizon. The paper from Sul et al. (2017) suggests that sentiment analysis have a bigger impact on small companies. Same line of results was attained by Bukovina (2015) that found that sentiment can influence stock price and volume of trade. Study from Duz Tan and Tas (2021) found that firm-specific *Twitter* sentiment contains information for predicting stock returns. The positive tone of *Twitter* sentiment is more pronounced in small and emerging market firms which is consistent with the literature stating that small firms are hard to value and emerging market firms contain high information asymmetry. Author Olof (2019) processed equity analysts reports through Bag-of-Words, term frequency-inverse document frequency, and Word2vec to label sentiment features. He found that analyst 'upgrade' or 'downgrade' recommendations were the most informative labels (neutral labels do not convey much information). Later, he applied logistic regression jointly with convolutional neural networks to classify the sentiment. Author Schlaubitz (2021) trained a DistilBERT model using a financial phrase databank to conduct sentiment analysis on Swiss earnings reports and news articles. The model predicted sentiment with an accuracy of 90%. The sentiment analysis discovered that a relatively low percentage of earnings reports expressed negative sentiment, even when companies reported lower earnings. In contrast, the news articles were more balanced between classes. Study concluded that predictive ability of text sentiment on future stock returns is weak as only 4 out of 15 companies under study showed significant connection under a linear regression analysis. Research has shown that sentiment within analyst reports can help explain price movements Suzuki et al. (2022). Rybinski (2020b) found that the forecasting ability of NLP sentiment improves when combined with traditional macroeconomic data. Corbet et al. (2015) examined the Greek and German markets, studying the impact of analyst recommendations across three market states: rising, falling, and crisis. They found that sell recommendations increase volatility, while the impact of buy recommendations varies depending on the market state.

Analysts have their own biases (Pursiainen, 2018; Li, 2022; Karmaziene, 2023) or

face limitations in time and resources, as noted by Le and Trinh (2022) and Kim et al. (2022). According to Thas Thaker et al. (2018), analyst reports account for 66% of the price movement. Similarly, Bandyopadhyay et al. (1995) demonstrated that long-term stock price changes are driven by profits (explaining 60% of price variation), while short-term stock price movements are influenced by the sentiment surrounding non-earnings factors (Nyakurukwa and Seetharam, 2023).

Presently, buy-side investors are exploring additional insights by analyzing the manner in which those words are expressed in the earnings calls with senior management following the company earnings releases. Technique was first employed to analyze the sentiment of Federal Reserve Chairman in the press conference to follow Federal Reserve Open Committee meetings where monetary policy is discussed Gordon and Lunsford (2024). The concept that audio recordings might offer insights into the genuine emotions of executives has captured the interest of some of the globe's most significant investors (Megaw, 2023).

Financial domain LLM StonkBert Pasch and Ehnes (2022) showed that the predictive capability of the model depends on the informational value of the underlying text data and that news sample outperformed both the blog as well as the annual reports sample. Firms predicted as 'good' by StonkBERT showed an average performance of 16.83% in the 12-month period after the prediction was made. The firms that were predicted as 'average' showed a performance of 4.72% and the 'bad' predictions of -3.17%. Experiments with PIXIU Xie et al. (2023) showed that the quality of the instructions rather than the model size is critical for the LLM performance. Models not fine-tuned with financial prediction datasets, show limited performance on stock movement prediction. Author van Binsbergen et al. (2021) found that FinBERT did not perform well when applied to hedge fund reports as 95% of those reports on short-selling included words as 'fraud' and 'misconduct' while Finbert was trained on companies' annual reports and conference calls with a language not mentioning those words. In study by Fatouros et al. (2023), ChatGPT exhibited approximately 35% enhanced performance in sentiment classification and a 36% higher correlation with market returns vs FinBERT in the short term. ChatGPT Yue et al. (2023) is a LLM developed by OpenAI. ChatGPT latest version is based on the GPT-4. It has been trained on a vast corpus of text data and can generate coherent and meaningful responses to a wide range of questions and prompts. ChatGPT-4 exhibits distinct behavior compared to the older GPT-3 model. ChatGPT-3 followed a text-in and text-out approach where it received a prompt string and provided a completion to be appended to the prompt. However, the ChatGPT-4 model operates on a conversation-in and message-out basis.

While integrating sentiment analysis and AI into financial markets offers numerous benefits, it also presents challenges such as algorithmic bias, data privacy issues, and the unpredictability of human emotions (Sahani, 2024).

2.3 Autonomous Financial Agents

2.3.1 The rise of Large Language Models

Although finance has embraced AI since its early stages (Bickley et al., 2022), the full potential of conversational AI technologies within this domain remains largely untapped (Yue et al., 2023).

The emergence of Generative Pre-trained Transformers (Liu et al., 2023c) marked a significant leap forward in Natural Language Processing. This technology allows Large Language Models to extract sentiment with far greater nuance and insight compared to traditional methods such as bag-of-words analysis.

Zhao et al. (2024) provides an overview of the applications of LLMs in finance, including automating financial report generation, forecasting market trends and analyzing investor sentiment. Li et al. (2023b) further developed the application of LLMs in finance providing a pathway to move from light models to customized LLMs. Krause (2023a) focused on three AI models: ChatGPT, Bard and Bing AI concluding with the need to address challenges such as contextual understanding, factual accuracy, bias, fairness and domain adaptation. The idea of domain specificity was further developed by Ling et al. (2023) who applied LLMs to solve sophisticated problems in specific domains and met many hurdles caused by the heterogeneity of domain data. The FinGPT model proposed by Liu et al. (2023b) provides an open-source foundational model in the financial domain that can later be customized through Low-Rank Adaptation. Guo et al. (2023) trained a model for the asset management industry.

2.3.2 Prompt engineering

A prompt is text containing a set of instructions that customizes, refines or improves the capabilities of a GPT model (Liu et al., 2023a). Prompt engineering is an emerging discipline that converges with artificial intelligence, computational linguistics and user experience design. It involves skillful creation of accurate and impactful inputs, termed prompts for LLMs. It consists of structuring prompts to shape the output of the LLMs. Liu et al. (2023a) classified the typology of the prompting methods.

Prompting enables interaction with the probability distribution of LLMs. In some cases, devising effective prompts can be challenging, leading to potentially inaccurate or contextually inappropriate responses that emphasize the importance of engineering instructions to achieve accurate results in different domains. These unexpected outcomes are commonly known as "hallucinations" (Rawte et al., 2023). LLM hallucination was mitigated by Peng et al. (2023) recurring to external sources which continually refine LLM prompts through an iterative process with the aim to enhance the model response. This refinement is informed by feedback derived from utility functions such as the factuality score of the responses generated by the LLM.

Writing effective prompts may require paying attention to the following characteristics (Velásquez-Henao et al., 2023):

- Clarity and precision
- Contextual information
- Desired format
- Verbosity control

Literature review 2020-2023 on prompt engineering/LLM spikes after the launch of ChatGPT as per 4Q23 (Figure 2.1). Of the 425 papers reviewed on prompt engineering during the period 2020-2023 there are hardly no mentions related to the financial domain or financial asset management (Wang et al., 2023b) (Figure 2.2).

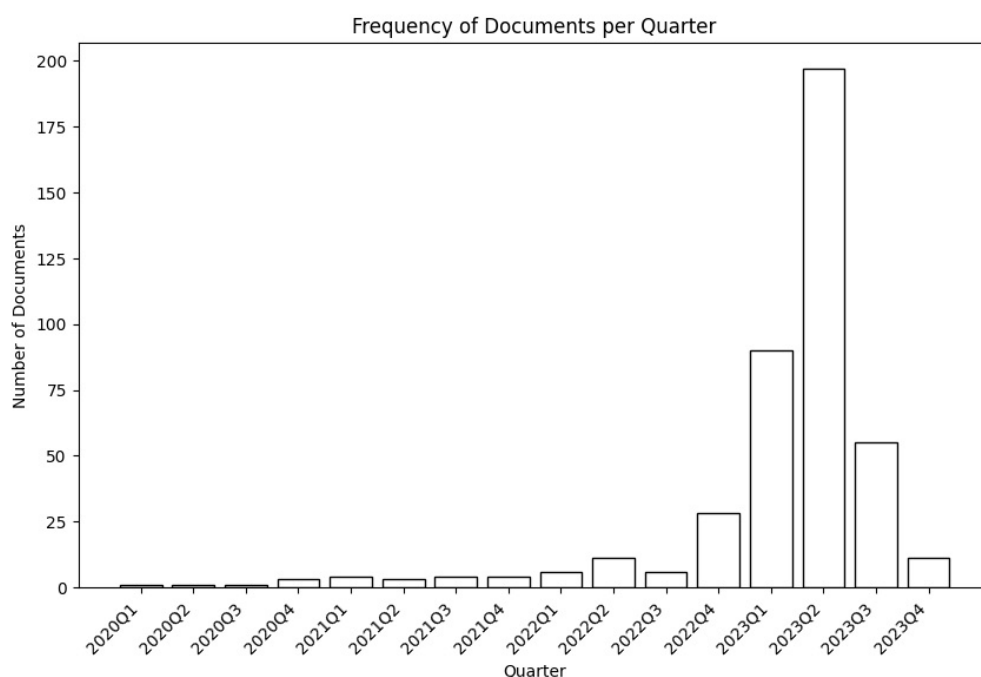


Figure 2.1: Prompt related papers evolution. www.promptingguide.com

Muktadir (2023) introduced the concept of an Automatic Prompt Engineer (APE), a method for automatic instruction generation and selection that outperforms human-generated prompts in 19 out of 24 NLP tasks. This is further supported by Clavié et al. (2023) that demonstrated the superior performance of a zero-shot GPT 3.5 turbo classifier with a well-designed prompt in job type classification. Feeding prompts into an LLM typically requires coding skills, which is why Arawjo et al. (2023) developed a graphical interface for model selection, prompt template design and hypothesis testing. Zhou et al. (2023) provided a historical overview of prompt engineering, emphasizing the evolution of techniques and the growing importance of ethical considerations in the field.

on extensive knowledge, CoK is initiated by formulating multiple provisional rationales and answers while simultaneously pinpointing pertinent domains of knowledge. In cases where there is no clear consensus among the answers, CoK systematically refines the rationales by integrating insights from the identified knowledge domains. These refined rationales serve as a more robust basis for the consolidation of answers. Unlike previous approaches that predominantly rely on unorganized data, CoK also harnesses structured knowledge to enhance its performance. White et al. (2023) provided prompt patterns as reusable solutions for solving common problems. Specifically in the financial domain, Krause (2023b) studies prompt techniques to improve financial analysis and investment decisions using ChatGPT, Google Bard and Bing AI chatbots.

2.3.3 Autonomous Financial Agents

AFAs essentially comprise different agents collaborating to accomplish a task. An agent, defined as an AI device with perception, cognition and action capabilities is proficient in processing data, making decisions and executing actions (Cheng et al., 2024). Being a state-of-the-art technology, there is still some confusion on the scope of the different terms used: "agents", "AI agents", "autonomous agents" or "LLM agents". AFAs represent a new way to interact with AI. We evolve from giving instructions to giving a broad goal just to see how to think about, how to solve the problem and execute a solution on its own. LLMs are central controllers for AFAs to obtain human-like decision-making. The interaction between an LLM and an AFA is similar to the interaction between the human brain and body parts (Lilian Weng, 2023b). Xi et al. (2023) conducted a survey on the rise and potential of LLM-based agents, identifying a lack of a general model for designing AI agents. Wang et al. (2023b) presented a comprehensive review of LLM-based autonomous agents (see Table 2.1). As per the literature, we can group the studies into four topics as follows:

Table 2.1: Autonomous agents taxonomy papers 2021-23

Social Science		Engineering		Natural Science	
Social simulation	11	Computer science	16	Science education	7
Psychology	4	Gral. auton. agents	14	Docu./ Data mgmt.	1
Political sci./Economy	2	CS/Gral. auton. agents	1	Sci. ed./Doc./Exp. assist.	1
Research assistant	1	Robotics/Embodied AI	19	Sci. ed./Exp. assist.	1
Jurisprudence	1	Industrial automation	3	Docu./ Exp. assist.	1
Psychology/Political sci.	1	Civil Engineering	1		

Cognitive modeling in AFAs

Lin et al. (2023) designed an agent inspired by the dual-process theory of human cognition, incorporating both a short-term, intuitive memory ("swift") and a long-term,

reflective memory ("sage"). This dual-memory approach allows the agent to handle immediate tasks swiftly while also reflecting on past experiences for better decision-making. Yu et al. (2024) developed FinMeM, an agent with a memory module that processes multi-source financial data and adapts to volatile market environments. This layered memory system aligns closely with human traders' information processing but extends beyond human cognitive limits, enhancing the agent's trading performance.

Collaborative frameworks

Hu et al. (2023) proposed a reinforcement learning-based mediator model that enables efficient and cost-effective interactions between algorithmic agents and LLMs. This model significantly reduces interaction costs and improves agent performance. Talebirad and Nadiri (2023) introduced a collaborative framework for LLMs, leveraging multi-agent systems to handle complex tasks more effectively. This framework addresses limitations and shows potential applications in various domains by enabling agents to work together seamlessly. Liu et al. (2023d) and Chen et al. (2023c) advanced the field by designing dynamic agent interaction environments and freeing agents' roles and capabilities to boost adaptability. This approach enables more flexible and efficient collaboration between agents. Hong et al. (2023) discussed the issue of the "cascade of hallucinations," which arises from the absence of "sanity checks" in agent interactions. According to Huang et al. (2024), a single-agent model is inadequate for handling complex financial scenarios. Another challenge in designing AFAs is the varying timeliness of different types of financial data. Luo et al. (2023) focused not on the structure of single-round prompts, but on the process of multiple prompt iterations from an optimal control perspective.

Prompting techniques

Li et al. (2023c) developed TradingGPT, which uses LLMs as the decision core for trading agents. By leveraging a layered memory system to integrate historical actions and market insights, TradingGPT provides better trading recommendations as it gathers more context and background information. Wang et al. (2023a) proposed the Learning through Communication (LtC) paradigm, a training approach that enables LLM agents to continuously improve through structured interactions with their environment and other agents. This method has demonstrated versatility and efficiency across diverse domains. Liu et al. (2023b) addressed the scarcity of available financial data by automating the collection and curation of real-time data from over 34 diverse sources. This initiative provides researchers with an accessible resource to develop their own LLMs.

Agent utilization

Guo et al. (2023) introduced Shai, an open-source foundational model continuously pre-trained and fine-tuned using a financially focused training corpus. Shai is capable of handling professional qualification exams, tailored tasks, open-ended question answering and risk assessments. Kolonin et al. (2023) developed the Adaptive Predictive Portfolio

Management Agent, which uses experiential learning and a "try and fail" methodology. This agent employs a swarm of subordinate agents in a virtual environment to identify optimal strategies and predict price movements using social media data. Zhang et al. (2024) developed a multi-modal foundational agent with tool augmentation for financial trading: text, numerical and visual inputs plus enhanced memory anchored on well-known trading strategies result on 36% profit vs benchmark on six financial datasets, including stocks and cryptocurrencies. Yang et al. (2024) breaks down input using prompting technique CoT, algorithm dynamically configures appropriate model application strategies from multi-source LLM. Huang and Tanaka (2022) introduced the concept of information-gathering and investment-decision agents for portfolio management that are not limited to a specific asset class and can be reused.

After carefully reviewing existing literature, we identified the research gap shown in Figure 2.3:

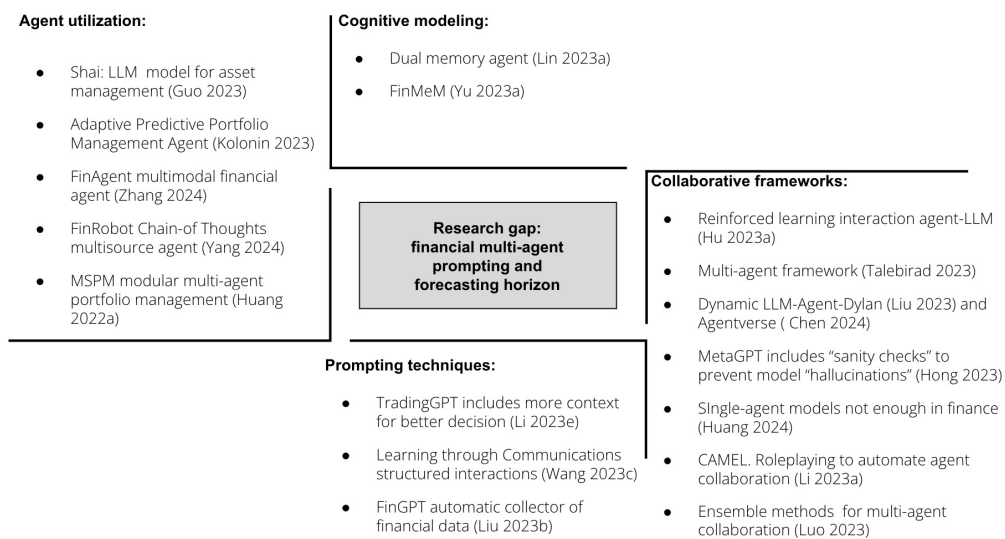


Figure 2.3: Multi-agent literature research gap

For the LLM based AFAs to perform the full cycle of tasks of asset management, namely client profiling, portfolio selection, performance evaluation, rebalancing and reporting, AFAs need specialized agents (Qian et al., 2023), (Hong et al., 2023) that can be trained through:

- Fine-tuning to improve models skills.
- Retrieval-Augmented Generation (RAG).

The field of AFAs is rapidly evolving with the potential to revolutionize the way individuals and institutions make financial decisions. AFAs are computer programs that can learn, reason and act independently to manage financial portfolios. They differ from robo-advisors in that AFAs are not limited to providing investment advice, they can

also execute trades, manage risk and provide personalized financial planning. AFAs are complex systems that use machine learning to learn from historical data and make real-time investment decisions. This context interaction is beyond the scope of robo-advisors. Other differences are:

- Transparency: Robo-advisors are generally more transparent about their investment strategies and fees than are AFAs.
- Customization: Robo-advisors typically offer a limited range of investment options, whereas AFAs may offer more customization.
- Regulation: Robo-advisors are subject to the same regulations as traditional financial advisors, whereas AFAs are still in the early stages of development and may not be as heavily regulated.
- Data: The quality of the data used to train the LLM is crucial to its performance.
- Cost: AFAs can be expensive to develop and maintain.

There are numerous recent developments regarding LLM-based multi-agents (Cheng et al., 2024): www.brainstormgpt.ai, www.flowiseai.com, www.vectorshift.ai, www.microsoft.github.io/ai etc...

Researchers (Wang et al., 2023b) have identified specific roles for AFAs in learning from the environment to evolve themselves like humans. The framework states the following modules:

- Profile: Identify the roles of agents. Being specific and detailed in the job description improves performing.
- Memory: Place the agent in a dynamic environment and memorize previous actions. AFAs possess a remarkable ability to store and retrieve information, enabling them to learn and adapt in real time while retaining knowledge over extended periods. This dual-pronged memory system, comprising short and long-term memories, serves as the foundation for contextual learning and decision-making capabilities
- Planning: Place the agent in a dynamic environment and plan future actions. They break down large tasks into smaller, more manageable steps. They also have the ability to learn from mistakes and improve their performance over time. This makes them powerful tools for solving problems and achieving their goals
- Action: translating the agent's decisions into specific outputs. LLM agents are capable of actively seeking additional information from external sources such as APIs or vector stores, based on their dynamic decision-making processes. By utilizing a range of tools and employing semantic searches and vector databases, AFAs can deliver precise and accurate responses that align with the findings of their external inquiries. This approach not only enhances the overall quality of LLM responses but also helps mitigate common issues associated with LLM technology such as inaccuracies and hallucinations.

2.4 Summary

This chapter has explored the existing body of research at the intersection of AI, sentiment analysis and financial market prediction. The literature review highlighted significant advancements in the field while also identifying gaps that justify the research objectives of this thesis:

- RQ1: Can sentiment help to predict stock price trends?
- RQ2: Does Deep Learning time-series stock prediction with price, technical indicators and sentiment outperform classical forecasting models?
- RQ3: How effective are autonomous financial agents in forecasting stock prices?

The review of AI applications in financial markets demonstrates that Machine Learning and Deep Learning models have progressively enhanced forecasting capabilities, particularly in handling non-linear relationships within financial data. LSTM models, Transformers and Attention-based models have shown superior performance in capturing time-dependent patterns. However, existing studies focus primarily on price and technical indicators, neglecting the potential role of textual sentiment in stock price movements.

Market sentiment, derived from investor emotions and public discourse, has been widely studied as a determinant of stock price fluctuations. Sentiment analysis using NLP techniques has demonstrated predictive potential, particularly when applied to earnings reports, financial news and social media discussions. However, a major limitation in the literature is the reliance on lexicon-based sentiment models, which fail to capture elusive financial language. Transformer based models, such as BERT and FinBERT, have recently emerged as superior alternatives but require further validation in real-market scenarios.

- RQ1: The success of AI in stock market forecasting suggests that incorporating sentiment analysis could further refine predictive models. However, there is limited empirical research quantifying the extent to which sentiment from analyst reports and earnings calls improves predictive accuracy, justifying the need for this study. Sentiment extracted from financial reports and analyst calls may contain valuable information not fully captured by traditional price-based models. However, studies suggest that different sentiment extraction techniques yield varying levels of accuracy, highlighting the need to test multiple models (i.e., FinBERT, GPT-4) in this research.
- RQ2: Existing studies confirm that Deep Learning methods generally outperform traditional statistical models, but the combination of technical, fundamental, and sentiment-based factors remains underexplored. This thesis seeks to evaluate whether a multi-source approach enhances prediction accuracy.
- RQ3: Sentiment analysis is a key component of AI-driven decision-making in financial markets. As AFAs evolve, their ability to process sentiment-based signals

will be crucial. However, the literature lacks empirical comparisons between rule-based robo-advisors and autonomous agents leveraging LLMs which this study aims to address. The superior capability of the AI agent to find resources (i.e., Web search, API calls, databases...) is a significant leap dealing with such an evolving environments as financial markets though their effectiveness compared to existing AI-based strategies remains underexplored. This research investigates whether AFAs incorporating sentiment, technical and fundamental analysis can outperform standard robo-advisors and human-driven strategies.

These findings set the stage for the empirical research presented in the following chapters, where advanced AI techniques will be applied to real-market data to test the hypotheses derived from the literature review.

Chapter 3

Experiment design

Building upon the gaps identified in the literature review in Chapter 2, this chapter outlines the experimental design used to evaluate the effectiveness of sentiment analysis and AI in stock market prediction. The experiments conducted in this study are structured to address the three key research questions posed in Chapter 1. By leveraging DL models, sentiment extraction techniques and AFAs, this chapter presents a systematic approach to testing how AI-driven methods can enhance financial market forecasting and portfolio management.

Each experiment is designed to test specific aspects of sentiment-based forecasting and AI-driven investment strategies:

- Experiment 1: Sentiment Analysis for Stock Prediction (IBEX 35 Index). Research Question addressed: RQ1 ("Can sentiment extracted from equity analysts' reports help predict stock price trends?") This experiment focuses on analyzing sentiment extracted from equity analysts' reports and its correlation with stock price movements. By applying LLMs such as FinBERT and GPT-based Transformers, the study aims to determine whether textual sentiment signals can enhance stock price predictions beyond traditional price-based indicators.
- Experiment 2: Multifactor stock prediction (NASDAQ Index). Research Question addressed: RQ2 ("Does Deep Learning time-series stock prediction with price, technical indicators and sentiment outperform classical forecasting models?"). This experiment integrates sentiment analysis with traditional stock prediction factors, including technical indicators and fundamental data. By comparing Deep Learning models (i.e., LSTM, Transformers) with classical time series forecasting methods (i.e., ARIMA), the study assesses whether incorporating sentiment leads to superior predictive accuracy.
- Experiment 3: Autonomous Financial Agents (AFAs) for Market Forecasting. Research Question addressed: RQ3 ("How effective are autonomous financial agents in forecasting stock prices?"). The final experiment evaluates the performance of

AI-powered financial agents in stock market forecasting. It compares the effectiveness of AFAs using real-time sentiment, technical indicators, and macroeconomic data against conventional robo-advisors and human-driven strategies. In addition, it examines the role of prompt engineering in optimizing AI-based investment decision making.

The integration of sentiment analysis and AI can improve stock market predictions by bridge the gap between quantitative data and investor psychology. Our study assesses the capabilities of sentiment analysis for stock market prediction alone or in conjunction with other characteristics.

Another avenue for exploration is the varying pace of AI adoption in relation to a country’s level of development. We compared the Spanish IBEX 35, which is heavily exposed to the LATAM market, with NASDAQ, a highly liquid index that hosts some of the world’s most advanced technology companies (Figure 3.1).

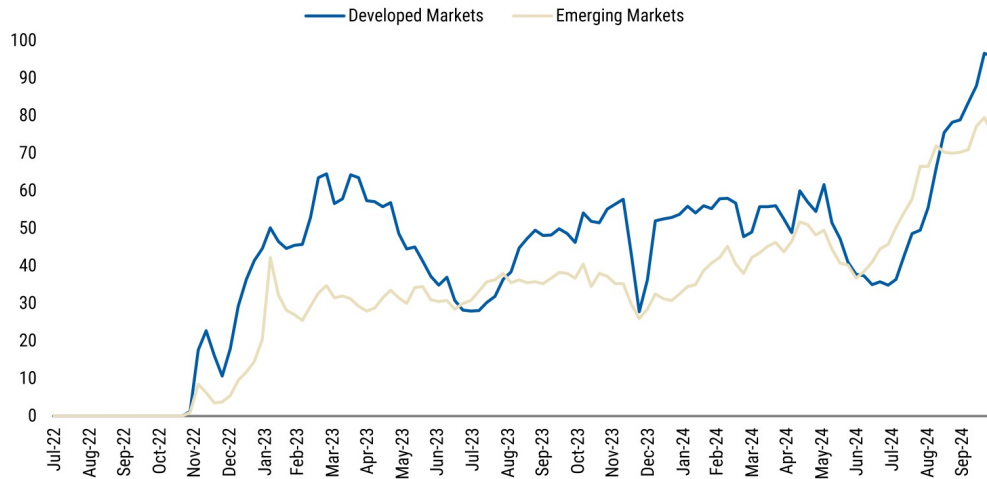


Figure 3.1: AI adoption according to geographies (Stanley et al., 2024)

The experiments in this chapter are designed to progressively build on each other, starting with isolated sentiment analysis (Experiment 1 in subchapter 3.1), advancing to integrated multifactor AI models (Experiment 2 in subchapter 3.2), and culminating in the deployment of autonomous AI-powered financial decision-making systems (Experiment 3 in subchapter 3.3). The results from these experiments will offer empirical insights into how AI and sentiment analysis can transform financial market prediction, tackling the key challenges highlighted in the literature.

The core methodology for the aforementioned experiments is based on Transformer architectures (i.e., BERT, FinBERT, GPT and LSTMs with Attention mechanisms) with their ability to process large volumes of unstructured financial text and extract meaningful sentiment signals. These architectures have revolutionized NLP and time series

forecasting. Transformers (Figure 3.2) were initially developed for sequential data modeling (i.e., text). Transformers utilize self-Attention mechanisms to identify relationships within data. Generative versions like GPT model data as a sequence and generate samples auto-regressively. They deliver state-of-the-art performance in text, image, and multimodal generation because of their scalability and ability to model long-range dependencies, although training them can be computationally costly, particularly with large datasets.

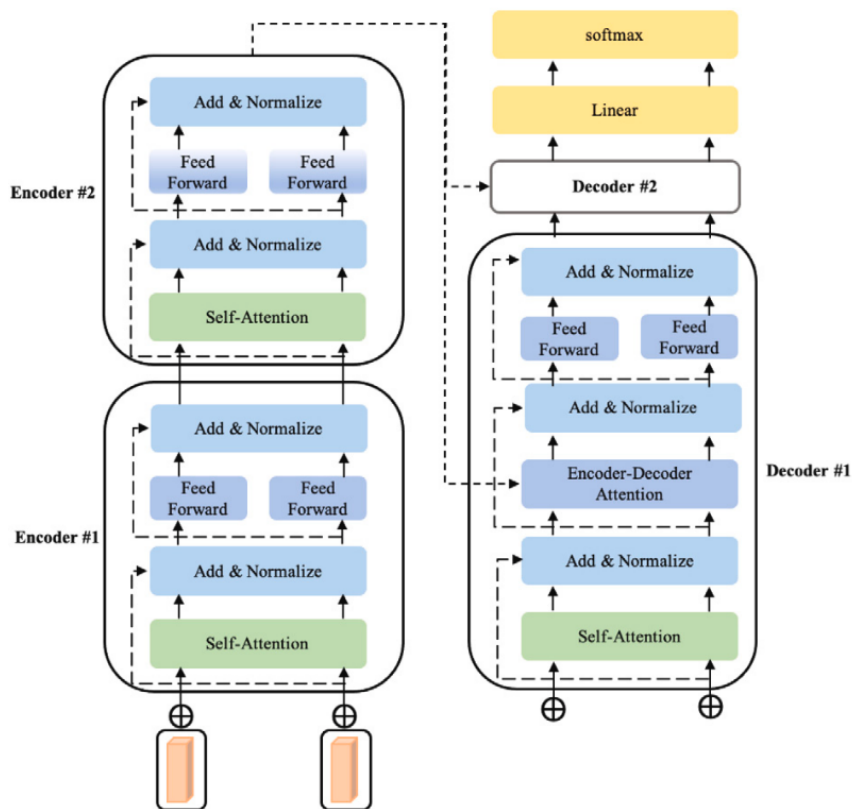


Figure 3.2: Transformer architecture (Zhang et al., 2022b)

This chapter bridges the gap between theoretical advances and empirical experimentation. The subsequent sections of this chapter detail the datasets, methodologies, and evaluation metrics employed in each experiment, ensuring a rigorous and replicable approach to assessing the impact of AI in financial forecasting.

3.1 Sentiment Analysis: IBEX Stocks Forecasting

3.1.1 Data

To answer the research questions, the sentiment embedded in the earnings call transcripts of the companies being studied was linked to their stock price movements to assess the forecasting ability of the language model. In this process, we compared the sentiment extracted through:

- RoBERTa.
- FinBERT.
- GPT-4.

Our corpus consist on sell-side research reports in English extracted from *Bloomberg* professional database for the period 2016-2022 for the 10 IBEX companies listed on Table 3.2.

We selected sell-side analyst reports vs buy-side research reports as per the following:

- Public availability: Sell-side reports are accessible, while buy-side reports are typically private.
- Market influence: Sell-side reports can significantly impact market prices and investor behavior due to their wide dissemination. Buy-side reports are proprietary.
- Research focus: Sell-side analysts provide detailed financial modeling, industry analysis and investment recommendations that are useful for understanding market trends. Buy-side reports are released for internal portfolio management.
- Regulatory standards: Sell-side reports are subject to strict regulations, ensuring transparency and reliability. Buy-side reports follow internal compliance guidelines.
- Consensus estimates: Sell-side analysts' estimates are aggregated to form widely tracked consensus benchmarks. Buy-side target price is not openly shared.

The 10-IBEX listed companies selected for this study comply with:

- Size of market/company (Tables 3.1, 3.2) as literature shows small company size is a factor in analysts' recommendations impact due to lower liquidity and less coverage (Lo, 2017).
- Sufficient sell-side coverage.

- Industry diversification. After implementation of MiFID II (Markets in Financial Instruments Directive II) (Keohane, 2018) that stated direct and explicit sell-side analyst payment, the number of reports available fell drastically: 15% as per Reuters (Reid, 2019), 47-53% as per CFA Institute (CFA Institute, 2019) and 30-40% according to European Commission (Risk Control Limited, 2020). They were free of charge until 2018.

Our sample is representative from the index as companies under study account for more than 50% index market capitalization covering different sectors despite the index banking concentration. That makes the analysis robust and non data dependent.

Table 3.1: European indexes size. October 2024 (Bloomberg)

Index	Market cap. (EUR Bn)	Country
FTSE 100	2.580	UK
CAC 40	2.410	France
DAX 30	1.840	Germany
IBEX 35	611	Spain

Our training data is unusual for LLM training in that it includes data from reliable sources rather than web scraped which is usually the case. The model entry is the text on the first page of sell-side equity research report which contains a detailed summary of the report. We performed preprocessing eliminating upper cases and any mention of the label assigned by the sell-side analyst in order to avoid forward looking bias. These labels basically refer to Sell/Underweight/Underperform (UP), Maintain/Neutral (N) or Buy/Overweight/Outperform (OP).

After extracting available research reports issued by Barclays (BAR), Deutsche Bank (DB), JP Morgan (JPM) and Credit Suisse (CS) provided by *Bloomberg* and discarding those that gave rise to errors, we studied a dataset of 379 reports (Tables 3.3, 3.4).

This dataset/corpus has been divided into two groups as follows:

- Training: 303 reports (80%)
- Testing: 76 reports (20%)

A bag-of-words analysis of the reports concludes the most frequent terms on the equity reports of the studied companies are shown in Table 3.5. Top 4 most-mentioned words have a frequency 18-26%.

As per analysis shown in Table 3.6, we noticed that for the 17 most mentioned words in Table 3.5, only 10 are from the financial domain (Loughran and McDonald, 2020). That gives an idea of the low discriminatory power of Bag-of-Words analysis (Amin et al., 2023).

Table 3.2: Selected IBEX 35 companies: IBEX share, volatility (Beta) and industry breakdown

Ticker	% Market cap.	Beta	Industry
IBE	13.96	0.62	Power generation
ITX	13.38	1.08	Apparel
SAN	11.63	1.32	Banking
AMS	5.84	1.28	Data processing
TEF	4.41	0.68	Telecoms
FER	4.33	0.94	Infrastructure
AENA	3.56	1.23	Transport
IAG	1.92	2.30	Airlines
ELE	1.63	0.63	Power generation
ACX	0.48	1.21	Steelmakers

Table 3.3: Sell-side equity reports for the selected IBEX companies issued by Barclays (BAR), Deutsche Bank (DB), JP Morgan (JPM), and Credit Suisse (CS)

Ticker	# docs	BAR	DB	JPM	CS
IBE	34	18	0	4	12
ITX	43	13	2	8	20
SAN	31	4	4	10	13
AMS	42	6	3	16	17
TEF	36	12	3	10	11
FER	37	2	0	35	0
AENA	43	11	2	12	18
IAG	32	9	2	6	15
ELE	39	20	0	7	12
ACX	28	1	8	6	13
Others	14	8	0	0	6
TOTAL	379	104	24	114	137

Table 3.4: Equity reports for the 10 selected IBEX 35 companies (2016–2022)

Ticker	2016	2017	2018	2019	2020	2021	2022	No. documents
IBE	3	2	2	2	9	7	9	34
ITX	0	0	0	6	12	15	10	43
SAN	0	0	0	0	1	15	15	31
AMS	0	0	4	6	20	8	4	42
TEF	0	0	0	4	9	10	13	36
FER	0	0	0	0	0	15	22	37
AENA	0	0	9	4	9	9	12	43
IAG	0	0	2	8	7	9	6	32
ELE	1	1	2	5	8	11	11	39
ACX	1	5	4	2	2	6	8	28
Other	2	2	1	1	1	3	4	14
TOTAL	7	10	24	38	78	108	114	379

Table 3.5: Top four most frequently used words in the analyst reports of the selected IBEX 35 companies, 2016–2022 (bag-of-words analysis)

Ticker	#1	#2	#3	#4	% top 4
IBE	EBITDA	Env. Issues	Wind	CAPEX	23.39
ITX	Revenue	Margins	Pricing	Leverage	22.31
SAN	Revenue	Equity	Dividends	Regulation	26.42
AMS	Revenue	EBITDA	Cash Flow	Margins	22.67
TEF	Revenue	EBITDA	Cash Flow	Competition	22.14
FER	Net Debt	Regulation	Cash Flow	EBITDA	17.78
AENA	Revenue	Cash Flow	EBITDA	Free Cash Flow	21.68
IAG	Revenue	Cash Flow	Margins	Free Cash Flow	20.32
ELE	EBITDA	Margins	CAPEX	Env. Issue	22.83
ACX	EBITDA	Steel	Net Debt	Inventory	21.00

Table 3.6: Frequency analysis of the terms extracted from 3.5

Word	# count
EBITDA	7
Revenues	6
Cash flow	5
Margins	4
CAPEX	2
Env. Issues	2
Free Cash Flow	2
Net Debt	2
Regulation	2
Competition	1
Dividends	1
Equity	1
Inventory	1
Leverage	1
Pricing	1
Steel	1
Wind	1

3.1.2 Methodology

The overall equity research sentiment is assessed for the report in its entirety, associating positive sentiment to upward movement (OP), neutral sentiment to neutral movement (N) and negative sentiment to downward movement (UP). Previous studies analyzing the sentiment of analyst reports have not made the distinction between current company situation and forward looking perspectives as they are closely intertwined. LLM understanding of the current situation is critical to assess sentiment going forward (Yang et al. (2020), Yang et al. (2023a)).

The size of the corpus of reports from different sectors plus the different time frames considered, smooth the potential noise of one-off effects: macro data, political events...

LLM categorized the sentiment of each report as positive, neutral, or negative, with positive sentiment indicating an upward trend (OP), neutral sentiment indicating no change (N), and negative sentiment indicating a downward trend (UP). The stock price movement is then analyzed from the day before each report's release across various future time windows: 2, 8, 16, 30 and 60 days.

The primary focus is on the degree of correlation between the sentiment of these reports and the stock price movements over different time horizons, considering various ranges of price changes $X \in [1\%, 3\%, 5\%, 7\%]$.

$Price_t$ is the share closing price the last day of the interval considered and $Price_0$ is the day before the report is released:

$$\frac{Price_t - Price_0}{Price_0} > X\% \rightarrow OP \quad (3.1)$$

$$\frac{Price_t - Price_0}{Price_0} < -X\% \rightarrow UP \quad (3.2)$$

$$[-X\%, +X\%] \rightarrow N \quad (3.3)$$

We frame a multiclass classification problem with three classes: Outperform (OP), Neutral (N) and Underperform (UP). We compare price evolution during the horizon selected $t \in [2, 8, 16, 30, 60]$ days for each range of price movement considered $X \in [1\%, 3\%, 5\%, 7\%]$ to qualify the movement as OP (Equation 3.1), UP (Equation 3.2) or NEUTRAL (Equation 3.3) if the movement is above X, below -X or inside +/-X.

The adopted F1 weighted metric can indeed help evaluate the model's performance by balancing the F1 scores across the different classes based on their support (i.e., the number of instances of each class). We have followed the four-step method depicted below, where the first two steps holds the practical implementation and the other two provide explanation and interpretation:

1. Calculating F1 score for each class:

- For each class (OP, N, and UP), compute Precision (Equation 3.4) and Recall (Equation 3.5):

$$Precision = \frac{TP}{TP + FP} \quad (3.4)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.5)$$

Where: TP: True Positive, FP: False Positive and FN: False Negative.

- Then, calculate the F1 score for each class using the formula (Equation 3.6):

$$F1 = \frac{2 \cdot (Precision + Recall)}{Precision \cdot Recall} \quad (3.6)$$

2. Weighting by support:

- In an F1 weighted metric, each class's F1 score is weighted by the number of instances (support) in that class. This means that classes with more samples will contribute more to the overall score.
- Formula for F1 weighted is (Equation 3.7):

$$F1_{weighted} = \frac{\sum_{c=1}^3 F1_c \cdot support_c}{\sum_{c=1}^3 support_c} \quad (3.7)$$

where $F1_c$ is the F1 score for each class c (OP, N, UP), and $support_c$ is the count of samples in that class.

3. Interpreting F1 weighted:

- The F1 weighted score provides an overall metric of the model's performance, taking into account both the performance on each class and the distribution of samples across classes.
- This is especially useful if the classes are imbalanced (i.e., more Neutral samples than Outperform or Underperform), as it prevents the metric from being overly influenced by a class with fewer samples.

4. Comparing Analyst vs. Model:

- By comparing the analyst's classification (ground truth) with the model's predictions using F1 weighted, we can get an indication of how well the model performs across all classes, accounting for any imbalances in the dataset.
- This metric will highlight if the model struggles with certain classes, or if it aligns closely with the analyst's judgments across all classes.

In summary, the adopted F1 weighted metrics becomes effective because it provides a balanced view of the model's performance across all categories (OP, N, UP) while adjusting for class imbalances. This ensures a comprehensive assessment of the model compared to the analyst's classifications.

Using transfer learning capabilities of BERT model (Devlin et al., 2018), we fine-tuned our models using a sample of our sell-side reports as training set to link sentiment and recommendation of analysts. With these fine-tuned models, we ingested text from the unseen analysts' reports test sample for the models to return a sentiment: positive, stock go up, negative, stock go down and neutral remains flat. Price was never used to fine-tune the model, so look forward bias is excluded (Sun et al., 2019). Methodology is similar to that of (Yang et al., 2023b) and (Kirtac and Germano, 2024).

RoBERTa

LLM model that stands for Robustly Optimized BERT Pretraining Approach (Liu et al., 2019). Pretrained on a massive corpus of text and code. It is based on Google's BERT model but makes several improvements including dynamic masking at each epoch, longer sequences, larger batches and Byte Pair Encoding (BPE), increasing the efficiency of handling text of RoBERTa vs BERT.

RoBERTa has been shown to outperform BERT on a variety of NLP tasks, including natural language inference, question answering and sentiment analysis. It is now one of the most popular and widely used Large Language Models in the world.

Numerous tests have been carried out with the model, being the one that has given the best results a training of 10 epochs, 16 of batch size and length limited to 512 tokens (sliding window).

FinBERT

LLM model that stands for Financial BERT (Araci, 2019; Yang et al., 2020). It is a pre-trained model with financial texts in English. FinBERT is a specialized language model designed for financial text analysis. This LLM is fine-tuned specifically for understanding financial language and concepts. It has been trained on a large corpus of financial documents, making it adept at tasks such as sentiment analysis, entity recognition and other financial NLP tasks. FinBERT is particularly valuable for applications in the finance industry, including stock market sentiment analysis, financial news sentiment tracking as it achieved the state-of-the-art on sentiment scoring (Financial PhraseBank).

We tested both RoBERTa and FinBERT using an sliding window technique as some texts may exceed their 512 token capabilities.

GPT

Since its release in November 2022, ChatGPT has revolutionized NLP. GPT-4, the most recent version at the writing time is even larger and stronger (Liu et al., 2023c) than GPT 3.5.

In the context of stock portfolio management (Ko and Lee, 2023), ChatGPT can be used to analyze market trends, provide insights into specific stocks or sectors and answer questions related to investment strategy. ChatGPT could generate a response that includes information on the company’s financials, recent news and market trends.

GPT-4 training set is publicly available web based information. Analyst reports are client-only data, not available on the web which prevents forward-looking bias.

We performed inferences on ChatGPT-4 with the objective of categorizing each of the reports. We upload the analysts reports through the ChatGPT-4 API and asked ChatGPT-4 to assign the above mentioned labels: UP, N or OP.

3.2 Sentiment Jointly with Other Features: NASDAQ Stocks Forecasting

3.2.1 Data

The analysis is conducted on the top 15 NASDAQ-listed stocks, based on market capitalization as of December 2020, as shown in Table 3.7. We monitored these stocks from January 1, 2015, to May 30, 2024, to predict their daily opening prices over a forecast horizon spanning multiple days.

The NASDAQ index was chosen because it represents a liquid market under robust supervision that helps prevent market manipulation. Additionally, it includes significant components, primarily from the technology sector, which generate a large share of the *tweets* related to stocks (Mao et al., 2012), forming the basis for the hypothesis being tested. The top 15 companies on the NASDAQ make up around 61% of the total market capitalization of the index. In contrast, the top 15 companies in the S&P 500 Index account for about 33% of the index’s total market capitalization. Since the *tweets* and news are specific to individual companies, the larger the market share of the top 15 companies, the more accurately the sentiment can forecast the overall behavior of the index.

The data used in this research consisted of historical daily opening prices of shares, along with technical and sentiment indicators, as provided by the *Bloomberg* professional platform. To account for any trend effects, we adjusted the daily opening prices of the stocks accordingly.

In addition to the open price and volume of the stock, we took into account the sentiment on Twitter and the news for each of the top 15 NASDAQ-listed companies.

Feature engineering or feature extraction refers to the process of transforming unstructured raw data into a vector space. This crucial phase requires expertise in the subject matter. The more problems feature extraction solves, the less challenges the machine learning system faces (Long et al., 2019). This implies that the use of machine learning becomes simpler and the better we are able to organize our input data from the large amount of information available, because the data features used to train the machine learning models have a huge influence on the performance they can achieve.

Sometimes, including irrelevant features in the input data can decrease the accuracy of many models and increase running time (Hira and Gillies, 2015).

Then, pre-processing of sentiment information was carried out by adding four new columns to the dataset, positive and negative ratio for Twitter/X and news sentiment analysis. For example, the positive tweet ratio was calculated as the count of positive publications divided by the total number of publications.

The second part is the calculation of the volatility. In this case, the method used to calculate the volatility of the stock has been the standard deviation of the daily percentage change in the price of the stock. Although volatility is not a raw data column, it is one of the most important metrics in market data. Volatility is used in a wide range of formulas, systems, and applications in the field of financial engineering and quantitative finance (Bhowmik and Wang, 2020; Chowdhury et al., 2022). After calculating volatility, the next step is to determine the trend. In this project, both the price trend and the volume trend are calculated. The inclusion of volume is based on its frequent use by traders as a metric to predict market movements (Bordino et al., 2012). The following step involves removing the *NaN* values, as the percentage of *NaN*s relative to the total data is minimal, making their removal appropriate.

- NASDAQ stocks:

Table 3.7: Top 15 NASDAQ index components by market capitalization

Ticker	Company	Index (%)
AAPL	Apple Inc.	11.38
MSFT	Microsoft Corp.	10.55
AMZN	Amazon.com Inc.	7.43
GOOG	Alphabet Inc. Class C	4.12
FB	Facebook Inc. Class A	3.96
GOOGL	Alphabet Inc. Class A	3.80
TSLA	Tesla	3.73
NVDA	Nvidia	3.63
PYPL	Paypal	2.25
ADBE	Adobe	2.16
CMCSA	Comcast	1.91
CSCO	Cisco Systems Inc.	1.72
NFLX	Netflix	1.70
PEP	PepsiCo Inc.	1.53
INTC	Intel Corp.	1.47

For these stocks, we retrieve daily Open, High, Low, Close prices, Volume and *Twitter* & news sentiment plus RSI_{14d} technical indicator (Dash and Dash, 2016). We used open prices because most of the returns are realized overnight with an

open gap the next day. We calculate a trend component for price (Eq. 3.8) and volume (Eq. 3.9):

$$P_x^{Trend} = 2 * \frac{(P_x^{Last} - P_x^{Open})}{(P_x^{Open} + P_x^{Last})} \quad (3.8)$$

$$P_x^{VolTrend} = P_x^{Trend} * Volume \quad (3.9)$$

- Investor sentiment for those stocks are assessed through Twitter & News sentiment provided by *Bloomberg* professional platform:

- Twitter sentiment. Since its launch in March 2006, Twitter has grown to become one of the most widely used social media in the world, with 353.9 million monthly users in 2023 (www.oberlo.com/blog/twitter-statistics). *Bloomberg* began to monitor and categorized *tweets* beginning 2015 through is highly regarded *Bloomberg Labs* group that consistently ranks top in the annual NLP competition *SemEval* (SemEval: The 18th International Workshop on Semantic Evaluation, 2024). The most mentioned companies on Twitter are mainly those that comprise the FAANG (Facebook, Apple, Amazon, Netflix, and Google).

Souza et al. (2015) defined equation (Eq. 3.10). He calculated the relative sentiment of a company by counting how many positive ($G(t)$) and negative ($B(t)$) messages were received that day t about that company:

$$S_R(t) = \frac{G(t) - B(t)}{G(t) + B(t)} \in [-1, 1] \quad (3.10)$$

$S_R(t_0) = +1$, represents a day t_0 with the highest positive sentiment for the company considered. In contrast, $S_R(t_0) = -1$ indicates the highest negative sentiment, while neutrality is achieved when $S_R(t_0) = 0$

- News sentiment. As stock market participants change the way news is ingested, the tracking of event-driven news with machine reading of news is on the rise. Machines can read the news and act more quickly than a person. A 600-word piece would normally take three minutes to read because the average adult can only read 200 words at a time. These sentiment scores on a news item or *tweet* as important for a specific stock vary from -1 for negative things to 1, with 0 being neutral. When used in trading, sentiment can serve as a directional indication to determine whether stocks in your portfolio should be held long or short. A common behavioral presumption is that if a company receives good news, its share price will increase, and vice versa.

However, positive news leads to a quick increase in stock returns, while negative news results in a delayed response over a longer period (Heston and Sinha, 2016).

Regarding the portfolio construction, different configurations have been considered, including discrete periods of 15, 30, 60 and 90 days, in coherence with the stock prediction. In addition, configurations were set for three different categories of stocks, although all of them are related to technology. The first category deals with companies producing goods and services for people’s everyday working (AAPL, ADBE, MSFT, GOOG, GOOGL); the second category focuses on services for people and entertainment (AMZN, FB, PEP, CMCSA, NFLX, PYPL), and finally the third one looks at hardware-centered companies (CSCO, INTC, NVDA, TSLA). That is,

$$B = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \quad (3.11)$$

Such a structure in categories is just to build an example and it does make any limitation to the proposal made in the thesis. Starting portfolio life on Nov. 1st the investment with a horizon of 90 days. The design was forced to cover at least two categories and three for some experiments. The weight factor between covariance and the expected return per unit of investment has been explored in a logarithmic shape.

3.2.2 Methodology

Construction of a portfolio involves two key processes: asset selection and weight allocation. According to *Markowitz’s* theory, including assets with negative correlations can enhance risk diversification (Löf et al., 2023). To determine the allocation sizes for the different assets, minimum-variance and mean-variance models are used as reference approaches (Clarke et al., 2006; Goktas and Duran, 2019). They are based on the compensation strategy that provides a degree of stability to the portfolio over time. Almost all approaches to optimizing a portfolio of stocks deal with these two coupled problems.

The primary contribution of this research, instead of focusing on selecting individual assets, is to broaden the concept of targeted assets to encompass categories based on factors such as sector or other considerations like ESG performance. Then, the idea is to consider the variance to decide about weights, but in combination with the expected performance of these stocks, with some estimation of the computational uncertainty of the prediction while a minimum number of categories are required to build the portfolio. This aspect is very relevant since the selection not only looks for diversification like the variance principle, but also looks at considering the expected performance, which is more proactive (see Figure 3.3).

The proposed approach has the additional effect of enabling an uncoupled configuration, since the prediction of future evolution of stocks can be carried out at any time and with different models and techniques.

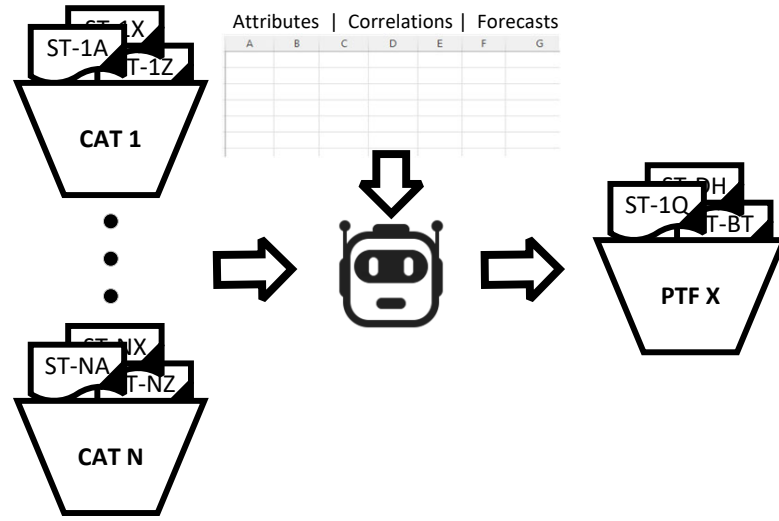


Figure 3.3: Selection of portfolio assets considering categories, variance and expected performance.

The optimization of portfolios built over families of stocks can be independently assessed, and when necessary, a rebalancing proposal can be the result of an optimization process that also considered transaction costs (see Figure 3.4).

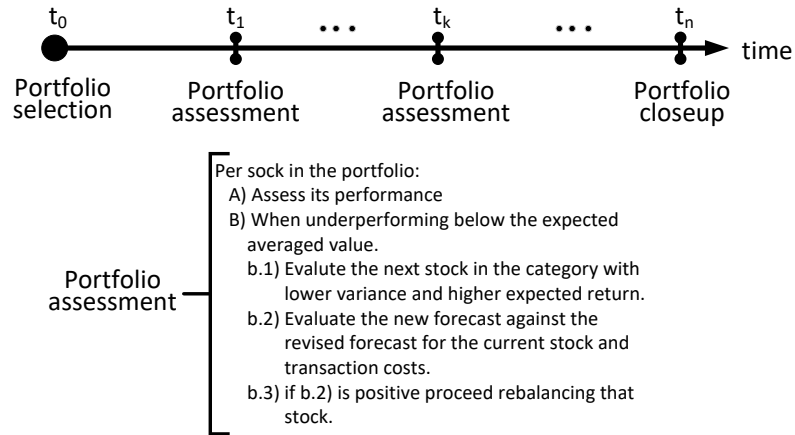


Figure 3.4: Discrete rebalancing strategy.

The optimization problem involves different subproblems. The first is related to the selection of stocks from different categories to set up the portfolio mix. The second is the rebalancing approach at discrete times according to the investment strategy selected by the owner of the portfolio. Finally, the last one, which can be done regularly and uncoupled from the two others is the regular forecast of the different stocks potentially targeted by the owner of the portfolio, since they are part of the different categories

under consideration. Let us formulate these problems.

The owner of the portfolio has selected the interesting groups of stocks according to his priorities, and each group is hereinafter named a category that depends on the owner's criterion. In the following, it is assumed that there are K categories. That is, let S be the set of assets, where $\{S = \cup_{k=1}^K C_k\}$. Obviously, different assets belong to different categories C_k . Formally, let us introduce the matrix B to describe the relationship between assets and categories, such as $B_{ij} = \begin{cases} 1 & \text{if asset } i \text{ belongs to } C_j \\ 0 & \text{otherwise} \end{cases}$

The selection of assets looks is based on minimum-variance and maximize the expected return in the defined horizon of portfolio. Finally, it takes the sum of the assets' weights to be one as a constraint and aims to find the optimal weight for the defined objectives.

$$\begin{aligned} \min_{\vec{\omega}} & \left(\vec{\omega}^t \sum_{cov} \vec{\omega} - \lambda \cdot \sum_{i=1}^{\text{card}(S)} \omega_i r(i, t_n) \right) \\ \text{s.t.} & \begin{cases} \vec{\omega}^t \cdot \vec{1} = 1 \\ r(i, t_n) = 100 \cdot \ln \left(\frac{P_i(t_n)}{P_i(t_0)} \right) \quad \forall i \in 1, \dots, \text{card}(S) \\ |B\vec{\omega}|_0 \geq N \end{cases} \end{aligned} \quad (3.12)$$

where Σ_{cov} represents the covariance matrix of the returns on assets when assets belong to different categories and one otherwise. $\text{card}(S)$ is the cardinality of the portfolio (number of different assets); and $P_i(t)$ represents the closing price of the asset i at time t . T represents the horizon for the planned investment portfolio, and N is the minimum acceptable diversification in categories of the portfolio measured by the norm L_0 . Finally, $\lambda > 0$ is a parameter that allows one to accommodate the relevance of variance and the range of benefits.

If we consider the rebalancing subproblem, the formulation at time t_s is presented in Eq. 3.13.

$$\begin{aligned} \min_{\vec{\omega}^j} & \left(\vec{\omega}^{jt} \sum_{cov} \vec{\omega}^j - \lambda \cdot \sum_{i=1}^{\text{card}(S)} (\omega_i^j r(i, t_n) - TC_i^j(t_s)) \right) \\ \text{s.t.} & \begin{cases} \vec{\omega}^{jt} \cdot \vec{1} = 1 \\ |B\vec{\omega}^j|_0 \geq N \end{cases} \end{aligned} \quad (3.13)$$

where ω_j refers to the weight of the asset i being replaced in the portfolio by the asset j at time t_s where this asset still belongs to the same category k as i . The $\vec{\omega}^j$ represents the whole set of portfolio weights where the asset i was replaced by the asset j , both belonging to the same category, and $TC_i^j(t_s)$ represents the transition costs for the replacement of the asset j by the asset i at time t_s . ϵ represents the minimum threshold to avoid taking actions without enough motivation.

Finally, the previous algorithms are supported by the need for an independent and robust estimate of the closing prices of assets at the selected time horizons $t_s \in t_0, \dots, t_n$. To facilitate this approach, this research was intended to use cutting-edge technologies from the machine learning area. As discussed in the state of the art, the most promising tools are related to the field of recurrent neural networks by using LSTM as well as to the Transformers, where multi-heading Attention mechanisms can help to estimate trends based on the observation of historical evolution observed in the past. In this case, the problem can be operated at the stock level, after observing that a single general model performed worse than individual models per asset Candemir and Karahan (2024)).

Therefore, additional effort has been made to model and understand the behaviors of these models to improve the estimation of $r(t_s)$. In 2017, a seminal paper Vaswani et al. (2017) introduced the Attention mechanism in LSTM, applied to the NLP domain. On the other hand, one of the first studies to suggest the use of an LSTM + Attention mechanism for multivariate prediction of time series is (Shih et al., 2019). Li et al. (2018) applied a multi-input Attention LSTM to distinguish valuable information from negatively correlated elements and remove their noise. Qiu et al. (2020) used an Attention mechanism to de-noise historical stock data. Similarly, Li et al. (2022) combined a Transformer encoder and an Attention mechanism together with social media *tweets* to predict stock movements. Social media are used solely to identify the most relevant stocks for the study.

This experiment uses neural network *Keras* front end run on Google's *Tensorflow* library using *Python* language to perform network implementation. To evaluate the predictive power of sentiment, we consider a persistence model as a benchmark and compare their results with LSTM, stacked-LSTM, bidirectional-LSTM, CNN-LSTM, Attention-LSTM and multivariate LSTM. We consider the predictive power through different forecasting horizons. Machine Learning algorithms with memory features extend the forecasting capacity in a daily-weekly range according to Shah et al. (2019) works. We have experimented with different numbers of neurons as part of our hyperparameter tuning process.

The aim was to find the optimal number of neurons that would result in a prediction model that is neither under- or over-fitting the data. Bhandari et al. (2022) found that 100-50-20 neurons may be the most appropriate configuration for a 3-layer model, which is consistent with our work. According to Zhang et al. (2018a), the number of hidden layers that minimize model error is 3. The different models use a sliding window of length 15, 25, 50 & 75 days to forecast 2, 8, 16, 30, 60 and 90 days ahead. MSE error of the models of the proposed models is calculated in the test set. All simulations have been repeated ten times, gauging the random set-up of weights and the random seed. The learning ability of the LSTM network is determined by the amount of neurons it contains. The batch size indicates how many input samples your LSTM should examine before changing the weights. Our LSTM univariate model involves 25, 50 & 75 hidden neurons in the hidden layer with a batch size of 500. Training the LSTM network involves selecting the training parameters: different window sizes: 15, 25, & 50 days. The activation function is ReLu and the optimizer selected is ADAM, while the mean

square error (MSE) was selected as a loss function. To produce a significant number of tests with such a diverse set of configurations, a tool has been developed and made available on gitgub <https://github.com/jordieres/Finance-AI>.

Unlike RNN, LSTM can preserve memory and state related to past activation rather than completely replacing it. They can retain characteristics for a long time thanks to the memory effect, which also makes it possible for back-propagation to occur across a number of constrained nonlinearities, reducing the risk that the gradient will vanish and allowing the RNN to learn the long-range dependencies across time steps. The rate of adjustments determines the gradient value.

Its structure is well known:

- The input gate regulates the extent to which the new cell state should be preserved.
- The forget gate determines the proportion of current memory to be discarded.
- The output gate regulates the amount of cell state that should be exposed to the higher levels of the network.

In addition, to contribute to the analysis of the DL capabilities, another technique has been implemented. It is Transformers, which are a type of neural network architecture introduced in 2017 by Vaswani et al. (2017). They revolutionized the field of NLP by providing a novel approach to sequence-to-sequence tasks. Unlike traditional RNNs, Transformers rely solely on self-Attention mechanisms to process input sequences in parallel, making them faster and more efficient. In connection with some other works (Vaswani et al., 2017; Tang and Matteson, 2021; Zeng et al., 2023) we wanted to evaluate their potential contribution to forecast stocks. The Transformer architecture uses multi-head Attention, which allows the model to jointly attend to information from different representation subspaces at different positions. This enables the model to capture complex contextual relationships between different parts of the input sequence. The other two sub-problems are handled by algorithms for conventional constrained minimization strategies.

The innovative idea here was to use a mechanism depending on the size of the window used to recall relevant memory length to estimate trends, enabling a variable normalization, which makes the process rather independent from monotonic tendencies and still keeps the meaning on derivatives as relevant for estimating trends. In fact, several authors have pointed out the relevance of human opinion for stock values (Yang et al., 2022; Valle-Cruz et al., 2022; Swathi et al., 2022), therefore data from social networks related to companies were included, as well as other potentially relevant variables, to allow for multidimensional time series forecast.

3.3 AI Agent Forecast Based on Sentiment, Technical and Fundamental Analysis: NASDAQ Stocks Forecasting

This study is about creating an AFA simulating a daily Investment Committee (IC) focused on stocks known as Magnificent Seven (MAG-7) for short term trading (Figure 3.5). The investment universe is the NASDAQ-listed companies called MAG-7: Apple, Microsoft, Nvidia, Tesla, Meta, Alphabet and Amazon for the period 17-30 December, 2024 and 23 December-3 January, 2025. These stocks were selected owing to their high media coverage and increased web searches in 2024 (Figure 3.6).

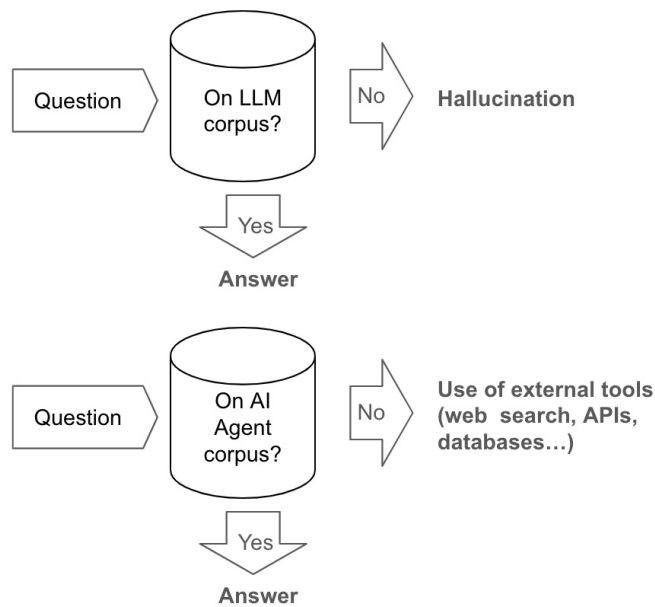


Figure 3.5: AI Agents vs LLM

The dataset includes hourly opening, high, low and closing prices for MAG-7. AFA will calculate the technical indicators itself.

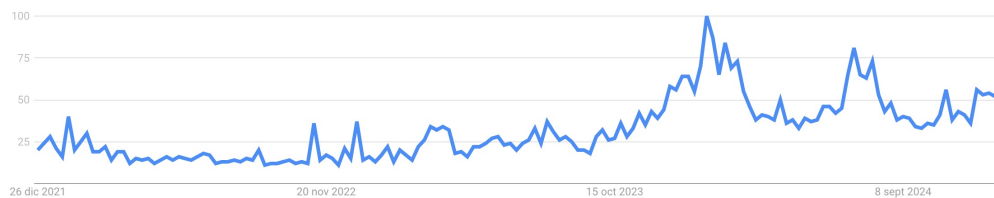


Figure 3.6: Google searches on MAG-7 2022-2024

Our study, test the accuracy of the forecasting based on LSTM and LLMs in $d+3$, $d+5$, $d+7$ and $d+10$ horizons vs time-series model ARIMA. We selected a short-term forecasting horizon as news sentiment is reflected in stock price in a range of 1-4 days (Dangl and Salbrechter, 2021). As daily news may be noisy, we increased the forecasting interval up to 10 days to smooth their one-off effect (Heston and Sinha, 2016). LSTM was trained with 2024 data on 10-epoch and 32 batch (increased to 64 in second trial). ARIMA baseline was fit with Python's `auto_arima`. We selected statistical ARIMA model as benchmark as it tends to outperform Deep Learning LSTM on a wide range of forecasting windows related to NASDAQ stocks prediction (Kobiela et al., 2022). The aim of this study is to conduct quantitative research but to reach a qualitative conclusion on whether LLMs integration capabilities of numerical and textual data outperforms pure numerical methods in stock market times series prediction. LLMs are probabilistic methods subject to a continuous training.

Our experiment consisted of building a digital twin for an IC (Figure 3.7). IC members were:

- **Stock researcher:** Retrieves company fundamental data and analyzes price charts for candle chart patterns and other technical indicators (i.e., MACD, RSI, etc) and generates buy/sell signals to identify trading opportunities based on price movements and trends. Analyst attempts to forecast short-term price trends.
- **Stock news sentiment researcher:** Analyzes news articles for sentiment (positive, negative, neutral) for MAG-7 companies and sectors.

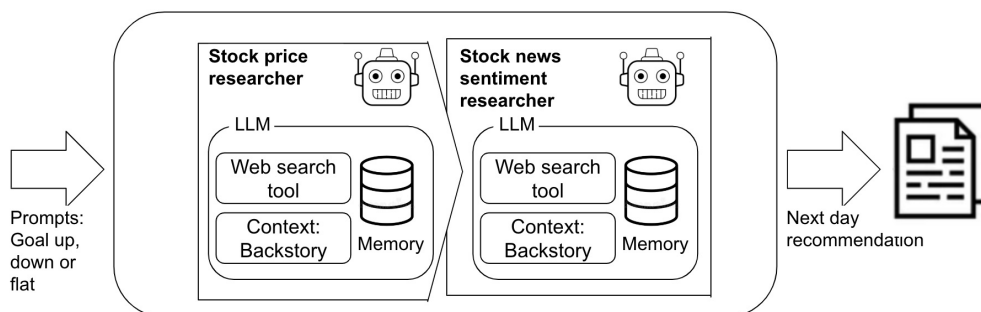


Figure 3.7: Multi-agent framework

We experimented with three prompting scenarios (Evdokimov, 2024):

1. **Fundamental+Technical+News:** It blends quantitative data (i.e., earnings, technical indicators) with qualitative data (news sentiment analysis).
2. **Fundamental+Technical:** Focuses on quantitative fundamental data and technical indicators (i.e., MACD, RSI...).

3. News: Analyzes company and industry news sentiment.

Stock news analysts fetch data from public sources (i.e., Google, Yahoo Finance, Investing.com, Insider Monkey, Seeking Alpha, MSN.com, etc...). The LLM engine in charge of evaluating news sentiment and price evolution was Anthropic's Claude Sonnet 3.5 (Morris et al., 2023).

We assigned our LLM agent three roles: stock price retrieval, stock technical and fundamental analysis and stock news sentiment analysis. The daily closing price retrieval from *Yahoo Finance* was performed through the Python's `yfinance` function. Technical, fundamental and sentiment analysis were performed based on following LLM prompts:

- a) Technical Analysis:
 - Identify key trends, support, and resistance levels from the stock graph.
 - Calculate and interpret important technical indicators (i.e., Moving Averages, RSI, MACD...).
- b) Fundamental Analysis:
 - Summarize recent financial performance and any notable events or announcements.
 - Assess the company's position within its industry and relative to competitors.
- c) News Impact:
 - Analyze how recent news might affect the stock's short-term and long-term performance.
 - Identify any upcoming events or potential catalysts that could impact the stock.
- d) Risk Assessment:
 - Evaluate potential risks and challenges facing the company and its stock.
 - Consider both company-specific and broader market/economic factors.
- e) Future Outlook:
 - Provide a detailed 3-month outlook for the stock, considering all analyzed factors.
 - Discuss potential scenarios that could significantly impact the stock's performance.
- f) Short-term Price Prediction:
 - Based on all available data, particularly recent news and

- market sentiment, predict the stock price for the next { X_days} trading days beginning {start_date}.
- Provide a range for each day (e.g., "Date: \$X – \$Y") and explain the reasoning behind each prediction.
- Provide a series of days and corresponding average of price range predicted (i.e., "Date: X").

As LLMs are probabilistic, they can generate slightly different outputs in consecutive runs in the same context. Consequently, the outputs of two runs of the same prompt version of the same model with identical data may occasionally vary. Several runs were run for each configuration.

3.3.1 Fundamental and technical investment analyst

Candlestick analysis originated in Japan in the 1700s (Nison, 1991). It is the oldest technical analysis technique and provides an idea of selling and buying balance in the market (Lin et al., 2021).

Although popular timeframes for candlestick analysis are 2-day patterns (Lu et al., 2012), high-frequency trading by machines has reduced this timeframe. Robo-trading is on the rise, with some estimates of 60-75% of the volume in the US market (Trading System, 2024).

The LLMs engine employed in this study has time-series forecasting capabilities by encoding time series as a string of numerical digits and solving the problem as a next-token prediction in text. Gruver et al. (2023) found that LLMs such as GPT-4 can surprisingly zero-shot extrapolate time series at a level comparable to or exceeding the performance of trained purpose-built time-series models. Claude works similar to other LLMs regarding analysis of time-series (Das et al., 2023).

3.3.2 News investment analyst

Price evolution is the confluence of buyers and sellers in which economic decisions are driven by expectations. The Efficient Market Hypothesis (Fama, 1970) states that available information is already reflected in the price and investors are rational.

The news sentiment agent is connected to the internet to browse relevant daily news through Python’s `duckduckgo_search` library. The LLM qualifies the sentiment and potential impact on a specific stock under study of the MAG-7 group. LLMs outperforms state-of-the-art sentiment extraction tools as per benchmarks on the competition SemEval-2017, where it improved the F-1 score by 22% (Kheiri and Karimi, 2023).

Chapter 4

Results

In Chapter 3, we designed three experiments to assess whether combining sentiment analysis with advanced AI techniques can improve stock market prediction. Chapter 4 now presents the empirical results of these experiments, highlighting how sentiment-driven insights and AI models performed in practice. This chapter is organized in three parts corresponding to the experiments (addressing RQ1, RQ2 and RQ3) and the key findings from each are summarized below:

4.1 Sentiment Analysis: IBEX Stocks Forecasting

The sentiment analysis of equity analysts' reports revealed that the majority of the reports carried a neutral sentiment with fewer negative sentiments (UP) and a considerable number of OP reports which is consistent with the incentives mentioned earlier (Grant et al., 2015). This suggests that analysts may exhibit a degree of optimism in their reports. This trend is evidenced by examining the proportion of predicted OP/Buy labels in the test group, where both LLM extracted sentiment and analysts' target assign very low UP recommendations (Table 4.1).

Table 4.1: Analyst recommendation and LLM extracted sentiment (%)

	Analyst	RoBERTa	FinBERT	GPT
UP	21	17	14	21
N	49	53	50	45
OP	30	30	36	34
TOTAL	100	100	100	100

Once we extracted the report text sentiment:

- Checked the consistency of the label assigned by the author vs that extracted from the report text (Table 4.2). F1 weighted metric was used to check the coincidence between the sentiment tags assigned by the analyst and each of the 3 different models. Market capitalization, industry, volatility and analyst background were considered, being this last factor the most influential in the consistency of text sentiment and explicit recommendation.
- Considering different price range movements for different timeframes since the report issue date, we checked the forecasting capacity of the report recommendation vs the extracted label through the three LLMs (Table 4.7).

Analysts are not specially good at price movement prediction (Bradshaw et al., 2013) in any horizon as per the precision (Equation 3.4) shown in Tables 4.3, 4.4, 4.5, 4.6.

In short forecasting periods, variations in the stock market are usually small (less than 5%), which is generally considered a neutral position. The higher proportion of N labels for RoBERTa may be one of the reasons to dominate in the short term.

Table 4.2: Analyst vs LLM extracted recommendation (Equation 3.7)

	An. vs RoBERTa	An. vs FinBERT	An. vs GPT
IBE	1.00	1.00	1.00
ITX	0.89	0.94	1.00
SAN	1.00	1.00	1.00
AMS	1.00	0.95	0.85
TEF	0.92	0.64	0.92
FER	1.00	1.00	1.00
AENA	1.00	0.76	1.00
IAG	1.00	1.00	1.00
ELE	0.76	0.91	0.91
ACX	0.91	0.91	1.00

Over time, the stock market undergoes increasingly pronounced fluctuations, making it easier to identify instances of OP or UP. In these cases is where our fine-tuned RoBERTa model shows better performance as it tends to make estimates more focused on detecting upward or downward trends rather than adopting a neutral position.

4.2 Sentiment Jointly with Other Features: NASDAQ Stocks Forecasting

Since the Portfolio Optimization Problem deals with several subproblems and most of them rely on the stock forecast, the presentation of results will begin with this one.

Table 4.3: Stock price vs prediction
(Equation. 3.4). 1% price range

	2d	8d	16d	30d	60d		2d	8d	16d	30d	60d
Analysts	0.28	0.32	0.26	0.33	0.26	Analysts	0.43	0.37	0.30	0.36	0.32
FinBERT	0.28	0.30	0.25	0.30	0.24	FinBERT	0.45	0.39	0.32	0.36	0.30
RoBERTa	0.29	0.30	0.28	0.34	0.25	RoBERTa	0.47	0.42	0.32	0.37	0.30
GPT-4	0.25	0.30	0.30	0.37	0.28	GPT-4	0.39	0.34	0.29	0.39	0.30

Table 4.4: Stock price vs prediction
(Equation. 3.4). 3% price range

Table 4.5: Stock price vs prediction
(Equation. 3.4). 5% price range

	2d	8d	16d	30d	60d		2d	8d	16d	30d	60d
Analysts	0.45	0.45	0.45	0.36	0.33	Analysts	0.49	0.47	0.39	0.37	0.39
FinBERT	0.46	0.47	0.45	0.37	0.34	FinBERT	0.50	0.50	0.39	0.38	0.42
RoBERTa	0.49	0.50	0.46	0.39	0.38	RoBERTa	0.53	0.53	0.41	0.39	0.45
GPT-4	0.41	0.42	0.41	0.39	0.32	GPT-4	0.45	0.45	0.36	0.37	0.38

Table 4.6: Stock price vs prediction
(Equation. 3.4). 7% price range

Table 4.7: Comparison of stock price trend predictions by time horizon
(AN: Analyst, FI: FinBERT, RO: Roberta, GP: GPT-4)

		2d	8d	16d	30d	60d
1%	RO	AN	GP	GP	GP	
3%	RO	RO	FI-RO	GP	AN	
5%	RO	RO	RO	RO-GP	RO	
7%	RO	RO	RO	RO	RO	

As already presented previously both univariate and multivariate models have been explored. This thread starts with the data windowing process, creating the new feature matrix with the window size extracted from the configuration file. Then, the data are adapted to the ahead value for the different periods prediction. And finally starts the normalization process, in this case, the normalization method applied is the min-max normalization, both for univariate and multivariate cases.

After preprocessing, learning of the different models was undertaken using the cross-validation strategy. The error metric used was the mean squared error (MSE). For the Transformer model, the architecture featured an embedding layer, along with encoder layers that included a multi-head attention layer and a positional feed-forward network. Additionally, dropout was applied after the multi-head attention and feed-forward network to prevent overfitting. Dropout helps regularize the model by randomly deactivating neurons during training, enhancing its ability to generalize.

Different sets of models have been built as presented in Figure 4.4, where different configurations as parameters were tested to gain knowledge of the strengths and limitations of those techniques in the different assets.

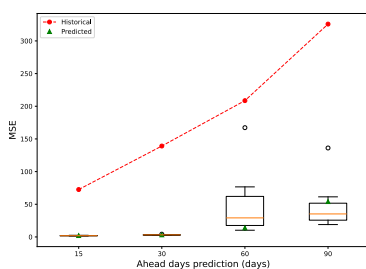


Figure 4.1: LSTM

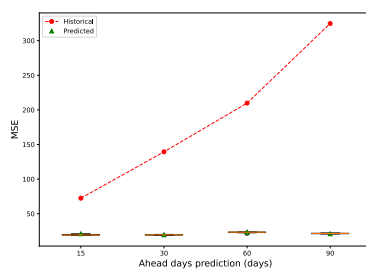


Figure 4.2: Transfor_1D

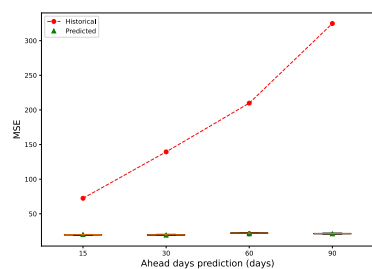


Figure 4.3: Transfor_4D

Figure 4.4: Performance comparison of different models using 10-fold cross-validation for GOOGLE class C assets

Where different architectures were explored, including different memory window lengths (2, 5, 10, 15 days) to analyze the influence of history on future evolution of the asset. In the case of LSTM, several configurations, such as pure LSTM, bidirectional-LSTM, stacked-LSTM, and Attention-LSTM, were explored. For the Transformer case, different models have been built, including a different number of transforming layers and number of Attention heads. In fact, both one-dimensional Transformers and multidimensional Transformers have been modeled, where different input variables have been considered in addition to the daily closing price(P_x^{Last}), such as P_x^{Trend} , RSI_{14d} , $P_x^{VolTrend}$.

Although the purpose of this thesis is not to show which configuration is the most convenient and also because behavior significantly depends on the stock itself, some rules have been derived. The Transformer technology with a high number of heads has good performance for long prediction periods (Figure 4.4, 4.5).

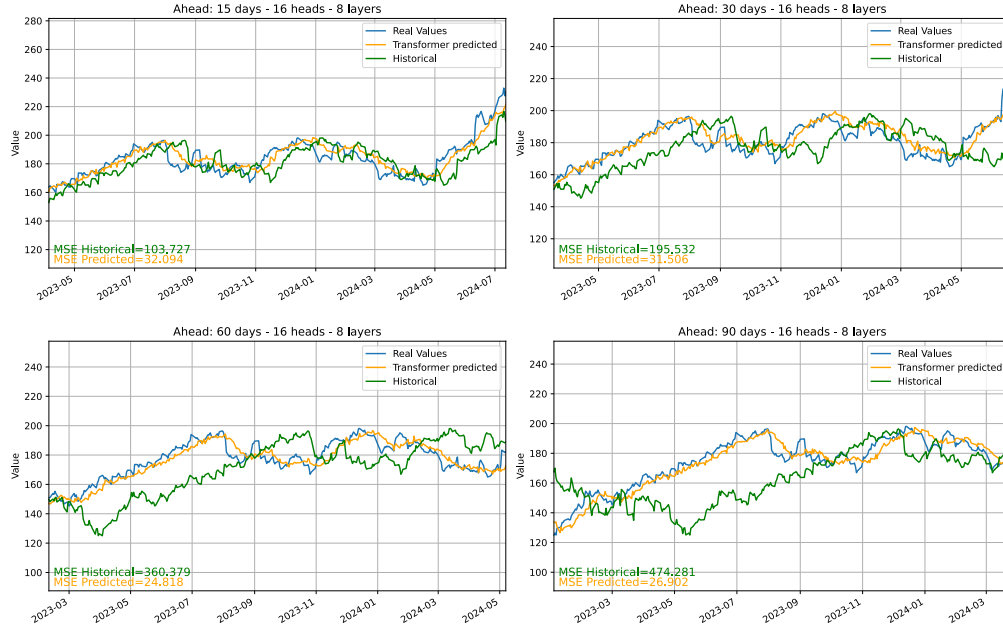


Figure 4.5: Prediction with different time windows for the asset value of Apple stock by using a multi-head Attention model with 16 heads and 8 layers

4.2.1 Portfolio performance evaluation

Sharpe ratio and drawdown have been selected for portfolio performance evaluation metrics, although several other metrics could also be adopted, such as turnover rate and break-even transaction cost.

Sharpe ratio (SR) measures the excess return per unit of risk. The larger the Sharpe ratio, the higher the risk premium per unit of risk. The Sharpe ratio is expressed as follows (Eq. 4.1):

$$SR = \frac{E(r_p) - r_f}{\sigma_p} \quad (4.1)$$

where $E(r_p)$ and σ_p are the average out-of-sample return and standard deviation of the portfolio, and r_f is the risk-free rate.

Drawdown (DD) reflects how much a portfolio's value falls from its historical peak. The drawdown is expressed as follows (Eq. 4.2):

$$DD_i(t_n) = \sup_{s \in [0, t_n]} \left(\vec{P}_i(s) - \vec{P}_i(t_n) \right) \quad (4.2)$$

where $\vec{P}_i(t) = \sum_{j \in Card(S)} (P_j(t) \cdot \omega_j)$ is the value of the portfolio i on day t , with $P_j(t)$ is the predicted price of the stock j at time t .

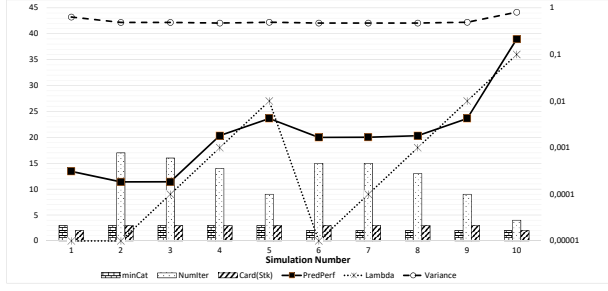


Figure 4.6: Variance vs λ

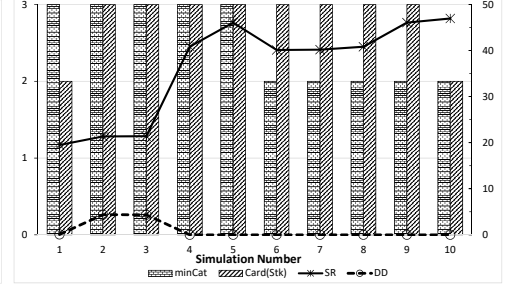


Figure 4.7: SR & DD factors

Figure 4.8: Alternative configurations for a portfolio. Each simulation corresponds to a different configuration of a portfolio. The simulation numbers correspond to the same experiment on both Figures 4.6 and 4.7

4.2.2 Portfolio evaluation

The portfolio optimization process is performed by solving equation 3.8, while rebalancing decisions are evaluated using equation 3.9. Although various strategies, such as Genetic Algorithms, can be employed for optimization, we opted for a numerical method in this small-scale experiment for simplicity. This method minimizes an objective function in a multidimensional space through a direct search, which is based on function comparison and is particularly useful for nonlinear optimization problems where derivatives are not readily available. However, the *Nelder–Mead* technique is a heuristic search method capable of converging to non-stationary points in problems that could be addressed by other methods (Gao and Han, 2012). Larger problems may require a tuning of the technique in use, depending of circumstances. However, grouping stocks into categories significantly reduces computational effort while increasing the conceptual meaning of investment decisions to be made.

The experiments carried out are summarized in Figure 4.8, where each simulation corresponds to a different configuration. In Figure 4.6 the bar graphs are related to the left scale, as well as the predicted performance. Meanwhile, the variance and λ curves are related to the right units, represented on a natural logarithmic scale. In Figure 4.7 the performance portfolio evaluation metrics have been presented, operating against the right scale of the graph, while the bar graphs reflect the structure of the portfolio in terms of the categories required and finally used and are related to the left scale of the graph. In this way, the interpretation becomes clearer for the decision-making process.

4.3 AI Agent Forecast Based on Sentiment, Technical and Fundamental Analysis: NASDAQ Stocks Forecasting

The results (Table 4.8) of the predictions for the periods December 16-27, 2024 and December 23-January 3, 2025 for the following prompting scenarios:

1. Fundamental+Technical+News
2. Fundamental+Technical
3. News

Table 4.8: Simulation results

Simulation dates	Scenario 1	Scenario 2	Scenario 3
Dec. 16-27, 2024	A.1, A.2, A.3, A.4	A.5, A.6, A.7, A.8	A.9, A.10, A.11, A.12
Dec. 23 - Jan. 3, 2025	B.1, B.2, B.3, B.4	B.5, B.6, B.7, B.8	B.9, B.10, B.11, B.12

A sample of a recommendation report is shown in Figure 4.9 and price recommendation statement is shown in Figure 4.10. LSTM and AFA forecast vs baseline ARIMA example is shown in Figure 4.11.

4.4 Summary

Across all three experiments, sentiment analysis proved to be an effective component of stock prediction strategies. The integration of textual sentiment (extracted via state-of-the-art NLP models) with AI-based predictive models enhanced forecasting accuracy and provided better guidance than using numerical data alone. Advanced Deep Learning architectures, especially Transformer models and Large Language Models, captured nuanced market signals from sentiment, leading to improved performance over traditional benchmarks. At the same time, the findings temper expectations by showing that these AI-enhanced strategies achieve only moderate success relative to market baselines. In conclusion, leveraging sentiment analysis in tandem with AI techniques offers measurable predictive benefits, while also highlighting the remaining challenges in achieving robust, above-market returns through these methods.

- Experiment 1 (RQ1: Sentiment Analysis. IBEX 35 case). The analysis of equity analysts' reports revealed a prevalent neutral-to-optimistic bias in recommendations, with relatively few negative (underperform) ratings. This aligns with known

Claude's Analysis:

Based on the provided stock price chart for Amazon (AMZN) over the last 241 days and the additional information, I'll conduct a thorough analysis of AMZN's stock performance and future outlook:

c) News Impact: Recent news highlights several positive factors for Amazon:

1. Analyst optimism: An analyst has reset Amazon's stock forecast, citing a potential "moonshot" opportunity. This suggests confidence in Amazon's future growth prospects.
2. Project Kuiper: Amazon's investment in satellite internet (Project Kuiper) through SpaceX launches demonstrates the company's commitment to expanding its technological reach.
3. AI focus: Amazon is frequently mentioned among top AI stocks, indicating its strong position in this growing field.
4. Hedge fund and analyst favorability: Multiple reports show that Amazon is viewed favorably by hedge funds and analysts as a top growth stock.
5. AWS strength: Despite some competitive pressures, Amazon Web Services (AWS) remains a formidable and profitable segment for the company.
6. Market cap milestone: Amazon's market cap has surpassed 2.4 trillion, with some analysts projecting it could reach 3 trillion in the future.

These news items generally paint a positive picture for Amazon's future, highlighting its diversified business model, technological leadership, and continued growth potential.

d) Risk Assessment: Potential risks facing Amazon include:

1. AWS competition: While AWS is strong, it faces increasing competition in the cloud computing space.
2. Regulatory scrutiny: As a tech giant, Amazon may face ongoing antitrust and regulatory challenges.
3. Economic sensitivity: Economic downturns could impact consumer spending and cloud adoption rates.
4. Labor issues: Recent news of worker walkouts highlights potential labor relations challenges.

e) Future Outlook: The 3-month outlook for Amazon appears positive based on several factors:

1. Strong upward trend: The stock chart shows a clear upward trajectory, especially in recent months.
2. Analyst optimism: Multiple analysts and hedge funds view Amazon favorably.
3. Diversified growth: Amazon's strength in e-commerce, cloud computing, and emerging technologies like AI position it well for continued growth.
4. Market leadership: Amazon's dominant position in multiple markets provides a strong foundation for future expansion.

However, investors should remain cautious of potential short-term volatility due to broader market conditions and the risks mentioned above.

Figure 4.9: Anthropic Claude output based on price and news

f) Short-term Price Prediction: Given the recent upward trend and positive sentiment, here's a speculative short-term price prediction for the next 5 trading days starting Monday, December 16, 2024:

Monday, December 16: 228—232 Tuesday, December 17: 230—234 Wednesday, December 18: 231—235 Thursday, December 19: 233—237 Friday, December 20: 234—238

Average of price ranges: Monday, December 16: 230 Tuesday, December 17: 232 Wednesday, December 18: 233 Thursday, December 19: 235 Friday, December 20: 236

This prediction assumes a continued positive trend but accounts for potential daily fluctuations. It's important to note that short-term stock predictions are inherently uncertain and subject to rapid changes based on news, market sentiment, and unforeseen events.

Figure 4.10: Anthropic Claude price forecast based on price and news

incentive structures for analysts to issue optimistic outlooks. When we extracted

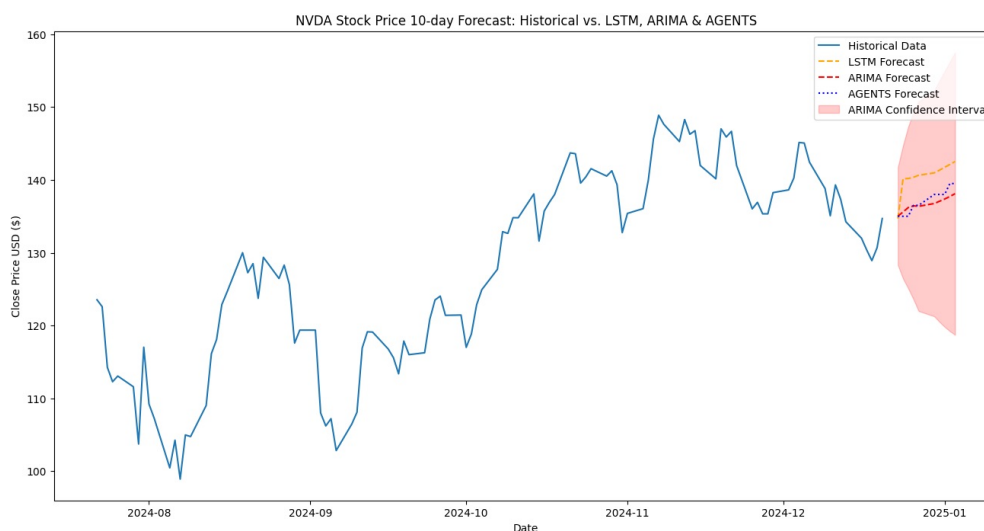


Figure 4.11: 10-day NVDA ARIMA, LSTM and AFAs forecast based on price technicals and fundamentals

textual sentiment from these reports using Large Language Models (FinBERT, RoBERTa, GPT-4), the sentiment labels generally mirrored the analysts' official recommendations, though individual analysts' writing styles caused some variance in consistency. Importantly, the predictive power of sentiment was modest but notable: neither the analysts' recommendations nor the extracted sentiments showed strong short-term prediction accuracy (precision scores were low across 2–60 day horizons). However, the fine-tuned RoBERTa model stood out by achieving the highest alignment with actual price movements, especially for larger stock price changes over longer horizons. This suggests that incorporating textual sentiment signals can marginally enhance the detection of significant market moves beyond what analysts' ratings alone achieve.

- Experiment 2 (RQ2: Multifactor AI Stock Prediction. NASDAQ case). This experiment evaluated whether integrating sentiment with technical and fundamental features in Deep Learning models yields better forecasts than classical time-series methods. The results confirm the effectiveness of the AI-based, sentiment-enriched approach. Deep Learning models (including LSTM variants and Transformers) that used price history, technical indicators, and sentiment outperformed a baseline ARIMA model in forecasting accuracy (lower MSE). Through extensive 10-fold cross-validation, we observed that model architecture plays a role in performance: for example, Transformer models with larger Attention head counts delivered superior accuracy for longer-term predictions. While short-horizon predictions saw comparable performance across architectures, the inclusion of sentiment features consistently improved the models' ability to anticipate price trends. In summary, AI models augmented with sentiment data achieved better predictive accuracy

than traditional models, underscoring that sentiment analysis provides valuable information complementary to purely numerical indicators.

- Experiment 3 (RQ3: Autonomous Financial Agents for Market Forecasting. Nasdaq case). The final experiment deployed an AFA that integrates real-time news sentiment, technical indicators, and fundamental data to simulate AI-driven investment decisions. The performance of the AFA was benchmarked in different information scenarios. The comprehensive AI agent (with sentiment, technical, and fundamental input) generated more informed forecasts and investment recommendations compared to scenarios excluding news sentiment. These results demonstrate that an AI agent leveraging sentiment can slightly outperform conventional models in portfolio performance metrics. However, the AFA's returns still largely tracked the broader market trends, reflecting only median performance overall. This indicates that while the inclusion of sentiment-driven analysis improves forecast quality and decision-making, current AI technology cannot reliably and consistently beat the market.

Chapter 5

Discussion

The growth of social networks and online forums has led to a significant increase in the volume of opinion data, requiring advanced technologies for real-time processing and analysis. AI, particularly through machine learning and Deep Learning Transformer models such as GPT and BERT, plays a key role in efficiently analyzing large datasets, identifying patterns, and forecasting market trends.

Research highlights the integration of sentiment data with historical market trends through AI for improved stock market prediction. Sentiment analysis, leveraging NLP techniques, extracts opinions from textual data sources like social media and news articles to gauge market sentiment. While promising, challenges remain, including algorithmic bias and the inherent unpredictability of human emotions in financial markets.

The findings of this research support the assertion that sentiment analysis, when integrated with traditional numerical data, offers enhanced forecasting capabilities in financial markets. In particular, the use of Large Language Models for sentiment extraction from news and analyst reports significantly improves the accuracy of stock price predictions when compared to traditional time-series models such as ARIMA and Deep Learning approaches like LSTM.

Several studies have compared statistical and Deep Learning forecasting approaches in financial markets (Nazareth and Ramana Reddy, 2023) with the following conclusions:

- ARIMA: Often performs well for linear, stationary time-series data but struggles with non-linear patterns or sudden shifts.
- LSTM: Excels in capturing non-linear dependencies and learning from sequential data but requires extensive training and is sensitive to hyper parameter tuning.
- AFAs: Shine in scenarios requiring data fusion: price trend and sentiment analysis.

From a methodological standpoint, the integration of textual sentiment from financial reports (via models like RoBERTa, FinBERT, and GPT-4) with numerical time-series data enhances the model's capacity to account for market sentiment shifts that often

drive price movements. As evidenced by the results, LLMs outperform traditional statistical models, particularly in volatile market conditions where external factors like investor sentiment play a pivotal role. This aligns with previous studies which argue that incorporating sentiment into forecasting frameworks can more accurately capture the impact of public sentiment and market news, offering a more robust and adaptive forecasting tool compared to relying solely on historical price data.

The interaction between sentiment and numerical data is particularly valuable in short-term forecasting (within 1-10 day horizons), where sentiment-driven market movements are typically more pronounced.

However, several limitations persist. While the performance of LLMs outpaces traditional models, challenges such as data quality, the potential for model overfitting, and the volatility of sentiment-driven predictions remain. Furthermore, the performance of LLMs can vary depending on the financial domain and market context, suggesting that fine-tuning models to specific sectors or regions might further improve forecasting accuracy.

5.1 Key Findings

5.1.1 Sentiment analysis: IBEX stocks forecasting

Language models provide a new approach to examining behavioral finance by enabling advanced sentiment analysis. This study analyzes the analyst reports of IBEX 35 companies over 2016-2022 to assess the consistency of the recommendations in analyst reports with their sentiment. Companies with higher market capitalization and fewer reports show the greatest alignment between sentiment and analyst recommendations regardless of the industry and volatility level. The consistency between analysts' explicit recommendation and the sentiment found by LLMs is notable for UK-based analysts. By time frame, we find that:

- Short term: The predictive ability of the RoBERTa model is marginally higher than that of the other two LLMs.
- Medium term: GPT-4 tends to outperform the other two models if markets stay flat; contrast, the RoBERTa model is more precise during larger price movements.

This study demonstrates the superiority of smaller but finance-related pretrained LLMs such as RoBERTa over larger general models such as GPT-4, which are not fine-tuned. This is similar to the findings of Xie et al. (2023). The predictive power of LLMs with no additional data is low, consistent with the results of Rybinski (2020a). LLMs pretrained in the finance domain such as FinBERT, but with no fine-tuning do not perform as well of other LLMs. Larger general models such as GPT-4 do not discriminate, as they tend to err on caution by using neutral labels that match with flattish markets (i.e., 1%-3%).

We also find that GPT-4 is not specifically trained on analyst reports but has more extensive generalist training, consistent with Gururangan et al. (2020). This may explain why RoBERTa better identifies trends over longer forecasting horizons.

Overall, while language models can provide valuable insights (Lopez-lira and Tang, 2023; Pelster and Val, 2024) and GPT-4 is suitable for summarizing vast amounts of information for retail investors (Kim et al., 2023a), they cannot replace the experience and knowledge of human investors. Nonetheless, by combining the insights generated by human experts' analysis with those from LLMs, investors can make more informed decisions, thereby reducing the potential for errors and biases (Cao et al., 2021). However, we find that the precision of NLP sentiment stock price trends hardly exceeds 50%, indicating its low predictive capability. These results are similar to those obtained by Rybinski (2020b).

Our research highlights the importance of the continued exploration and development of language models in the financial industry (Li et al., 2023a). Moreover, scholars should examine how such models affect market dynamics.

Our research contributes to the literature in the following ways:

- We show the consistency of analyst recommendations and potential bias, in line with Frijns and Garel (2021).
- We also show that the bigger the firm and the higher the analyst coverage, the better the GPT-4 forecast. These results are consistent with those of Li et al. (2023a).
- Crucially, our work is novel in its use of analyst reports as its dataset (i.e., research reports from IBEX 35-listed companies over 2016-2022). Similar studies have only covered the Korean stock market (Kim et al., 2023b; Cho et al., 2021).
- We also use a novel methodology, under which both domain-specific and general LLMs are used (FinBERT, RoBERTa, and GPT-4). Previous work on analyst reports based on ML has focused on discovering the importance of forecasting features (Sidogi et al., 2022).

However, some issues related to the development of LLMs need further attention:

- Secrecy and limited access to the training corpora of LLMs. Controversy continues about whether companies should open their proprietary LLMs to the public (i.e., *BloombergGPT* (Wu et al., 2023)), with OpenAI and META exhibiting opposite views¹. This remains a keystone in developing LLMs in the financial domain jointly with datasets such as The Pile Biderman et al. (2022), C4, and Wikipedia.
- Noise and instability. LLM strategies, especially when used by major institutional investors and hedge funds, may influence wider financial markets. Substantial

¹<https://www.nytimes.com/2023/05/18/technology/ai-meta-open-source.html>

capital allocation into specific securities and sectors because of LLM strategies can sway prices and market sentiment, potentially inciting herd behavior or unexpected repercussions that could engender systemic risks.

Overall, this study provides novel evidence on the predictive ability of state-of-the-art NLP models, such as FinBERT, RoBERTa, and GPT-4, using sell-side equity analysts' reports. Focusing on companies in the IBEX 35 index, this study demonstrates the potential of such NLP models to enhance the accuracy of stock price forecasts, even in markets with relatively high efficiency. Therefore, it is valuable for the literature on behavioral finance and ML applications in equity markets. This study also compares explicit recommendations by analysts with the implicit sentiment extracted by LLMs to examine the relationship between analyst sentiment and stock price dynamics. The results underline the limitations of human analysis by showing the complementary function of ML in financial decisions. Crucially, our work enriches the debate on the EMH by showing how analysts influence market efficiency through their sentiments and forecasts. Further, we provide a framework for further research on the intersection between NLP and financial markets. Practitioners such as investors and portfolio managers, can leverage NLP-derived sentiment analysis as a tool to refine decision-making processes, especially in anticipating price movements beyond analysts' explicit recommendations. Finally, our insights can help policymakers by showing the need to carefully consider sell-side analysts' role in influencing market dynamics, particularly in terms of ensuring transparency and mitigating biases in financial reports. Future research could consider more general applications of this approach to other markets, particularly emerging markets in which information asymmetry and imperfections are higher. Meanwhile, extending sentiment analysis by incorporating new alternative data sources, such as social media and macroeconomic indicators, may improve the predictive power. The main strength of this study, which shows the efficiency of using NLP models in analyst reports, is that it closes the gap between traditional financial analysis and advanced AI technologies. These findings contribute not only to the literature but also to actionable knowledge relevant to market participants striving for higher forecasting accuracy and optimized investment strategies.

5.1.2 Sentiment, technical and fundamental analysis: NASDAQ stocks forecasting

The approach adopted to establish the portfolio configuration facilitates the consideration of different AI-based algorithms. In this work, we have implemented several recurrent neural network-based algorithms, such as LSTM, stacked LSTM, CNN with LSTM and Attention LSTM. In addition, we have implemented different technologies, such as Transformers with both individual and multidimensional dimension and multiple Attention heads. The multidimensionality involved variables such as volatility, volume of operations, technical indicator RSI, sentiment for the asset from the general public and from specific operators, in addition to the future closing price. Such diversification in forecast algorithms brought additional insights, which evidenced that LSTM models have a better

performance for short-term prediction (around 30 days), while Transformers show greater stability in prediction. This behavior is not uniform across the assets, and every asset has its own behavior, which is an additional aspect where multiple modeling provides a good flexibility. When multidimensionality is considered, the improvement was lower than originally expected. The main reason for this behavior is the limited amount of data available, since we operate on a daily basis and, although we started collecting data when sentiment was initially recorded, the available data set includes nine years and a half, but when operating in such multidimensional space, the effect "curse of dimensionality" takes place (Poggio et al., 2017).

The different simulations represented in Figure 4.8 correspond to different mixtures of stock selections, except the first, which is a portfolio selected by humans (baseline) where no optimization was performed. As a result, the optimization process did not require any iterations. Simulations 2 through 5 were conducted under the condition that the portfolio must contain at least three categories. In each successive simulation, the weight (λ) assigned to the expected performance relative to covariance was increased tenfold. It is evident that for very low values of λ , the variance between assets remains low. However, as λ increases, the correlation between assets increases over time. This increase in correlation can reduce the resilience of the portfolio, although it also enhances the expected performance $r_i(tn)$.

The analysis of the information derived for the same simulations and shown in Figure 4.7 requires one to know that the bar charts are related to the left scale and the curves are related to the right scale. It makes clear that the SR factor for portfolio configurations with higher number of categories involved are less productive but more diverse, while configurations with two categories allow proposals involving three categories effectively involved but SR is significantly high. On the other hand, simulations two and three show that the Drawdown factor is positive, which means that the highest effectiveness of the portfolio can be reached earlier than its expected life.

The hybrid approach, which integrates asset covariance and expected performance with artificial intelligence techniques, emerges as a valuable tool. It offers a knowledge-based framework that is well suited for the portfolio configuration optimization technique employed in this study.

The ability to assess the portfolio across different time periods and re-optimize it as needed creates a robust environment for evaluating and adjusting the chosen strategy based on empirical evidence. This approach also allows estimation of uncertainties associated with the assets, providing an additional layer of information that can be used to mediate decision-making.

5.1.3 Autonomous Financial Agents

The inherent disparity between financial market data and language-based data precludes AFAs from becoming machine learning models capable of achieving autonomous training and self-sufficient trading abilities. Instead, AFAs primarily function as tools to augment human capabilities in performing the fundamental tasks essential for portfolio

management: feature identification, model hypothesis formulation, trading, backtesting and ongoing monitoring.

AFAs may struggle to interpret the qualitative or contextual information that affects stock performance such as news events, geopolitical factors or regulatory changes.

In a multi-agent system, roles combine their interests towards a common goal of eventually trading. In the real world, it is difficult to assess the motivations of different agents to trade. Advancements in reinforcement learning have heavily revolved around the ability to provide more comprehensive explanations.

It's important to note that while AFAs can provide valuable support, human expertise and judgment remain crucial in the financial industry. It is advisable to use AFAs as a complement to human decision-making rather than as a replacement (Cao et al., 2021).

AFAs outperform LSTM for short-term stock price prediction as of this study when the context involves non-time series data sources (i.e., sentiment, news) with forward looking context. The optimal features of AFAs is related to previous forecasting horizon. AFAs integrative capabilities of quantitative and qualitative data provides the highest accuracy in short-term price trend forecasting.

Other points to be noticed from the experiment:

- Structured prompt, results in structured answer.
- Incorporating news sentiment produces better predictions than price action alone which is consistent with Dahal et al. (2023).
- Sometimes the robot refers to not being an investment specialist to provide investment recommendations (i.e., META).
- LSTM forecast usually lies beyond ARIMA confidence interval. LSTM performs better with longer training periods. In order to set common conditions for LSTM and AFAs, we limited historical data to 2024. Results for the experiment show the superior capabilities of LLMs as zero-shot learners which is consistent with Gruver et al. (2023).
- META and TESLA are the stocks with the highest error compared to ARIMA baseline. AAPL, MSFT and GOOG forecast are the closest to that of ARIMA.
- Prompting scenario 1 outperforms scenario 3 in AFA forecasting accuracy compared to ARIMA. Scenario 2 the worst of the three. These results on both trials are consistent with the blending capabilities of AFAs to consider both hard data and soft data for an improved forecast.

5.2 Findings Implications

The superior performance of Transformer based models in sentiment analysis suggests a paradigm shift in financial forecasting methodologies. Traditional models like ARIMA, which rely on linear assumptions and historical data patterns, may not fully capture the

complex, non-linear relationships inherent in financial markets. In contrast, Transformer models, with their ability to process and interpret large volumes of unstructured data, including news articles, social media posts, and financial reports, can uncover nuanced market sentiments that influence asset prices.

Integrating sentiment analysis into forecasting models enables a more comprehensive understanding of market dynamics. By incorporating real-time public sentiment, financial institutions can enhance their predictive accuracy, leading to more informed investment decisions and improved risk management strategies.

The findings highlight the growing importance of AI integration within financial services. As AI technologies advance, their application in financial forecasting is expected to expand, offering more sophisticated tools for market analysis. Financial institutions may need to invest in AI infrastructure and expertise to leverage these technologies effectively, ensuring they remain competitive in an increasingly data-driven market environment.

Chapter 6

Conclusions and Future Research

This research substantiates the superiority of AI-driven models, specifically those utilizing LLMs for sentiment analysis, over traditional time-series models and Deep Learning approaches in financial forecasting. The integration of sentiment from analyst reports and market news offers a clear advantage in predicting short-term price movements, underscoring the potential of AI agents in transforming financial decision-making.

AI-driven forecasting agents, particularly those capable of sentiment analysis, provide more nuanced insights into market trends, surpassing the capabilities of traditional models like ARIMA and LSTM. These findings open the door for more sophisticated, real-time forecasting systems that can dynamically adapt to market conditions, offering investors a powerful tool to enhance portfolio management and investment strategies.

Future research should explore further optimization of these models by incorporating a wider variety of alternative data sources, including social media sentiment and macroeconomic indicators, to further refine forecasting accuracy. Additionally, continued investigation into the regulatory and ethical implications of using such AI tools in financial markets will be crucial as their adoption continues to grow.

6.1 Summary of Findings

This research has explored the application of sentiment analysis and AI techniques in stock market prediction and portfolio management. The key findings of this study are as follows:

1. Sentiment Analysis Models:
 - Sentiment from analyst reports and Stock Price Trends: The sentiment extracted from equity analyst reports using LLMs predicted stock price trends, particularly for short-term price movements. The results showed that sentiment analysis, as an indicator, offers value beyond traditional financial metrics in predicting stock price directions. Sentiment derived from earnings calls

and financial news also proved to enhance the prediction accuracy of stock price trends.

- Comparison of LLMs and traditional models: LLMs outperformed traditional lexicon-based sentiment analysis methods in terms of prediction accuracy and adaptability. While lexicon-based models are limited by predefined word lists, LLMs can capture the nuances and contextual meanings of financial language, offering more precise sentiment extraction and, consequently, more accurate predictions.

2. AI in Portfolio Management:

- Autonomous Financial Agents (AFAs): AFAs improved portfolio management strategies compared to traditional Robo-advisors. AI agents not only analyzed market data but also autonomously interacted with it, learning from past outcomes and refining their strategies over time. AFAs were able to adapt to changing market conditions in ways that rule-based Robo-advisors could not, resulting in more dynamic and robust portfolio management. AFAs excel in data fusion integrating price trends, sentiment analysis, and fundamental analysis unlike traditional models for stock prediction. These agents successfully bridge quantitative and qualitative data, making them particularly effective in capturing market shifts.
- Robo-advisors vs AI Agents: Robo-advisors, despite their efficiency and low-cost advantages, struggled with dynamic market shifts, especially in unpredictable conditions such as market crashes or sudden political events. In contrast, AI agents, which leverage Deep Learning and sentiment analysis, were more resilient and provided higher returns during volatile periods by adjusting strategies in real-time based on current data.

3. Effectiveness of Deep Learning Models:

- Performance of AI models: Deep Learning models, particularly LSTM, outperformed classical forecasting models like ARIMA in predicting stock prices. When combined with sentiment data extracted from financial news and social media, LSTM models showed a higher prediction accuracy, highlighting the potential of AI models that integrate sentiment analysis with traditional time-series forecasting techniques. Uncoupling the forecast of asset evolution from the optimization itself enables us to use different algorithms independently, indirect estimation for uncertainty can be derived and it can be incorporated as moderator in the selection of assets for the portfolio. In coherence with Behera et al. (2023), this thesis confirms the power of combining the mean value-at-risk (VaR) model and AI-based prediction of asset evolution for portfolio optimization. In fact, this thesis shows that technology supports not just the definition of the portfolio but also rebalancing during its life, including reconsideration of relative importance between individual criteria.

- Sentiment's role in Market Prediction: Sentiment analysis, especially when combined with price action and technical indicators, proved to be an essential factor in improving prediction accuracy. The integration of sentiment data from multiple sources such as news, analyst reports, and social media enhanced the AI models' ability to understand market psychology, thus improving their forecasting capabilities. LLMs improve short-term forecasting. RoBERTa demonstrates the best predictive accuracy, but with a narrow margin. Medium-term: GPT-4 performs better under stable market conditions, whereas RoBERTa is more precise in volatile scenarios

4. Implications for Financial Decision-Making:

- AI's impact on financial decision-making: The findings suggest that AI and sentiment analysis have a transformative impact on financial decision-making. AI agents capable of processing and learning from vast amounts of unstructured data -such as news, analyst reports, and earnings calls- enable more informed and adaptive financial strategies. These AI models are not only a tool for analysis but also an active participant in the decision-making process, continuously refining strategies based on new data.
- Improved risk management: By incorporating sentiment analysis into AI models, investors can potentially mitigate risk. AI agents were able to detect shifts in market sentiment before they were fully reflected in stock prices, providing a competitive advantage in risk management.

6.2 Contributions to the Field

Study provides insights for asset managers on leveraging AI tools for market forecasting and portfolio optimization and the development of profitable trading strategies for the short and medium term (Dingli and Fournier, 2017; Krollner et al., 2010).

The main contributions of this study are:

- Limited Application of LLMs in the financial sector despite the increasing popularity of LLMs in various disciplines, their adoption in the financial domain remains relatively scarce. This research aims to bridge this gap by exploring how language models can enhance stock market prediction.
- Evolution from robo-advisors to AFAs. The study investigates the transition from traditional robo-advisors to AI-powered AFAs, which not only provide investment recommendations but also execute trades and manage risks.
- Integration of sentiment analysis into predictive models. Most traditional stock market prediction models rely on time-series data and quantitative indicators. This thesis introduces an approach that incorporates sentiment analysis extracted from

analyst reports and financial news, adding a psychological layer of investor behavior that conventional models do not consider.

- Differentiation of AI models for financial analysis. The research explores how domain-specific financial models, such as FinBERT, compare with general-purpose models like GPT-4 in terms of accuracy and reliability for stock price prediction.
- Combination of technical, fundamental, and sentiment-based factors. The study argues that stock market prediction models can achieve superior performance by integrating technical analysis (price indicators), fundamental analysis (financial data) and sentiment analysis (NLP applied to news and reports). This holistic approach remains underexplored in the existing literature.
- Impact of LLMs on market stability. The thesis also examines how the increasing automation of financial decision-making through AI may influence market stability and investment dynamics.
- Ethical and market implications of AI in finance. AI models introduce potential biases based on training data, which can influence financial decision making. Regulatory frameworks need to evolve to address challenges related to AI transparency and risks of market manipulation.

In summary, this research fills a significant gap in the literature by integrating multiple advanced methodologies (NLP, LLMs, technical and fundamental analysis) into a framework that not only predicts stock market movements but also contextualizes them within investor psychology and the evolution of autonomous financial agents.

6.3 Limitations

Issues to be solved on the development of the LLMs portfolio management strategies are those of:

- Despite the successes of Deep Learning in computer vision, its application to time series models is still underdeveloped, with simpler autoregressive models often preferred over Deep Learning approaches (Jiang, 2021). Typically, factors such as interpretability, sparse data, usability, and training costs influence the adoption of these methods. While there is no one-size-fits-all solution to these challenges, deep models with Attention outperform traditional models in terms of overall performance, while also maintaining interpretability, and in many cases, they are even faster (Huang et al., 2019; Zhang et al., 2019).
- Secrecy and limited access to the training corpus of the LLMs (Figure 6.1). Controversy whether to open to the public their proprietary LLMs or not is still open with opposite views as those of OpenAI and META¹. *BloombergGPT* (Wu et al.,

¹<https://www.nytimes.com/2023/05/18/technology/ai-meta-open-source.html>

2023) will be a keystone on the development of LLMs on the finance domain jointly with datasets like The Pile Biderman et al. (2022), C4 and Wikipedia.

- Noise and instability. LLM strategies, especially when utilized by major institutional investors or hedge funds, hold the capacity to exert influence on the wider financial markets. Substantial allocations of capital into specific securities or sectors as a result of LLM strategies, can sway prices and market sentiment, potentially inciting herd behavior or unforeseen repercussions that can engender systemic risks.
- Environmental implications of energy consumption. The race for bigger and faster models imply the development of energy hungry AI chip processors. A debate on how to feed those datacenters is alive (Figure 6.2).

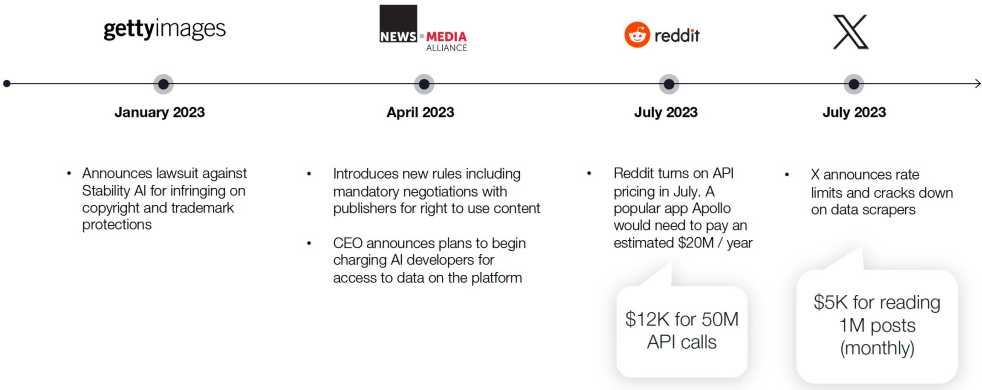


Figure 6.1: Data is a new currency (Kumar and Sharma, 2023)

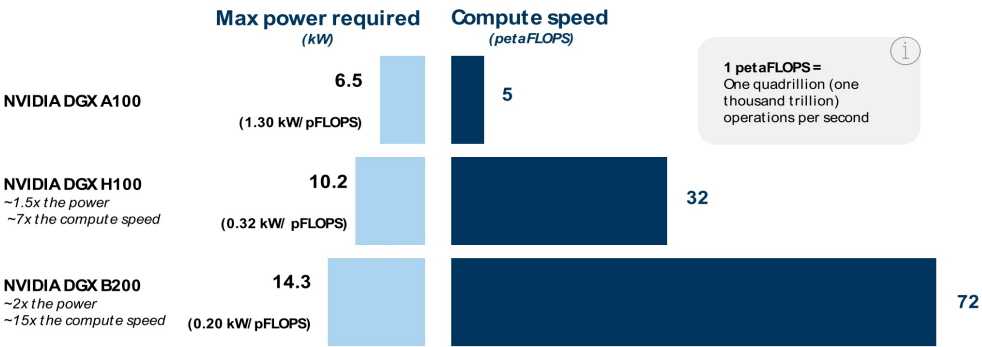


Figure 6.2: Processor capacity & related power consumption (Sheridan, 2025)

The number of roles in a AI agent prepared to forecast stock market can have a significant impact on the quality of the analysis but it is not a straightforward relationship as

there are many factors intertwined (i.e., macro-economy, fixed income, cryptocurrencies, alternative assets, etc)

The multi-agent configuration should be linked to the forecasting horizon. In a short-term horizon, including a macroeconomic analysts may provide little help as the impact of macro variables is medium to long-term. On the other hand, announcements or expectations are swiftly incorporated in the stock price (i.e., Federal Reserve Open Market Committee (FOMC) minutes or Central Bank Governor speeches) (Ehrmann, 2004). Testimonies during the first day of the Monetary Policy Report have the greatest impact, whereas forward guidance in the testimony explains only a small portion of the movement in stock prices. (Gordon and Lunsford, 2024). The short term effect of macroeconomic policies are taken care through the news investment analyst.

The benefits of increasing number of roles such as diversity and redundancy of information sources may get the collateral effects of information inconsistency and computational cost.

The right balance depend on the stock market idiosyncrasy (i.e., growth vs value stocks, volatility, regional coverage or inclusion of alternate assets like cryptocurrencies) and resources available (i.e., historical data, computational cost and trading frequency).

LLMs are not inherently designed for time-series forecasting. ARIMA and LSTM are purpose-built for capturing temporal dependencies which may give them an edge in purely historical data-driven predictions. On the other hand, AFAs require high-quality and diverse datasets to outperform ARIMA, LSTM models. Without these datasets, they might over-fit or misinterpret noisy signals. On what AFAs excel are in the transfer learning capabilities of their LLM engines and numerical and textual data blending capabilities.

Traditional models like ARIMA are more interpretable, providing clear relationships between variables. AFAs may act as black boxes, making their predictions harder to validate not to mention their superior computing cost.

6.4 Recommendations for Future Research

- Increase length of dataset as *Bloomberg* platform do include more data points. In our experiments we have used 1/2015-7/2021 daily data for 15 NASDAQ listed stocks with roughly 1600 samples in the dataset. A larger sample size for test and train datasets could not be gathered in this study as *Bloomberg* news and Twitter sentiment began January 1st, 2015. Enlargement of sample size will increase the accuracy and allow to include further input features without the risk of over-fitting, avoiding the curse of dimensionality.
- Incorporate more sentiment features (i.e., Google searches) to compute if opening price prediction improves.
- Size and volatility may be matter for further research to assess idiosyncratic stock forecasting behavior.

- Expanding the use of multi-modal LLMs and exploring the role of behavioral factors in financial AI systems.
- Compare the performance of different LLM models (Figure 6.3).

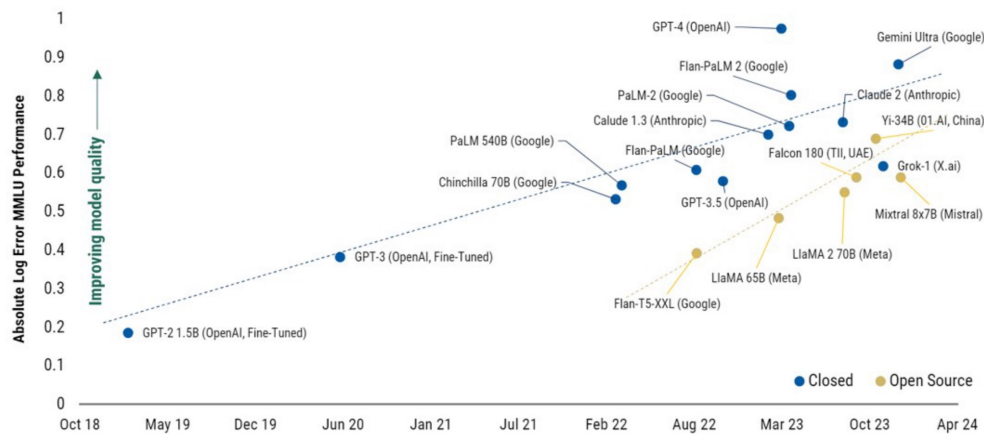


Figure 6.3: Model performance convergence: private & public (Stanley et al., 2024)

One point to consider is the use of a specific LLM devoted to time series analysis like Garza and Mergenthaler-Canseco (2023) for the technical analyst role. Another avenue for research may be to train an LLM specifically for candle chart pattern recognition. Hyper-parameter optimization is a key point in any research study and a methodology like that of Jiang and Li (2017) could be worth exploring.

While our study focuses on the comparative performance of Transformer models and ARIMA as benchmark, future research could explore the integration of other AI techniques, such as ensemble learning and reinforcement learning, to further enhance forecasting accuracy. Additionally, examining the impact of incorporating alternative data sources (i.e., Google searches, satellite images, credit card data...), could provide deeper insights into market behavior.

In conclusion, the integration of Transformer based sentiment analysis into AI agents represents a significant advancement in financial forecasting, offering more accurate and nuanced predictions compared to traditional econometric methods. This development has profound implications for financial analysis, AI integration strategies, and future research directions in the field.

Chapter 7

Practical implications

7.1 Stock Market Forecasting

7.1.1 Improved predictive performance based on sentiment

The research demonstrates significant gains in market forecast accuracy when modern AI models are used, particularly those integrating sentiment analysis. DL architectures, such as LSTM and Transformer models, consistently outperform traditional models like ARIMA in stock price prediction. Incorporating sentiment features further boosts predictive performance, uncovering insights not evident in historical price data alone. Sentiment analysis thus complements numerical data, enhancing model accuracy.

LLMs like FinBERT and GPT-4 outperform conventional lexicon-based methods by interpreting the nuanced language of finance, leading to more precise sentiment extraction and stronger forecasting capabilities. Firms aiming to leverage AI for market prediction should prioritize advanced NLP models capable of identifying subtle bullish or bearish signals that traditional methods may overlook.

A key takeaway is the validation of sentiment analysis as a critical tool for financial decision-making. Sentiment extracted from analyst reports, earnings calls, and news articles adds predictive value, especially for short-term movements, enriching traditional financial indicators. Capturing the "market sentiment" provides a more holistic view of market dynamics, supporting better-informed decisions. Both traders and AI systems can use sentiment indicators to optimize entry and exit points or manage risk by anticipating periods of extreme optimism or fear.

This thesis outlines practical methodologies for incorporating sentiment into investment processes. Advanced NLP models like FinBERT, RoBERTa, and GPT-4 can process real-time news, social media, and analyst reports, feeding sentiment scores directly into trading systems. This enables algorithms to account for the narratives driving price movements (i.e., pausing a buy order if negative sentiment suggests an unsustainable rally). Merging sentiment with technical and fundamental indicators leads to more robust strategies.

However, the research also notes that sentiment signals can be noisy and volatile. Sentiment analysis should therefore complement, not replace, traditional analysis and expert judgment. Used wisely, it adds a psychological dimension to financial forecasting, better aligning strategies with investor behavior. As AI adoption grows, sentiment analysis will likely become a standard part of investment and risk management, helping investors respond not just to market data, but to the emotions and expectations shaping it.

By incorporating sentiment data from unstructured sources (i.e., analyst reports, earnings calls, and social media) forecasting models gain a forward-looking perspective. LLMs like GPT-4 and FinBERT, combined with technical and fundamental indicators, outperform traditional models like ARIMA, marking a shift toward more adaptive and responsive financial forecasting. Rapid sentiment analysis allows decision-makers to react rapidly to market volatility, providing a competitive advantage.

7.1.2 Role of LLMs and prompt engineering

The use of LLMs such as GPT-4 introduces new considerations around prompt engineering. The study found that the structure and clarity of the prompts significantly affected the quality of AI agent decisions. Specifically, structured queries yielded more coherent and actionable responses. The testing of various prompting strategies revealed notable differences in forecast accuracy, with the most effective prompts combining quantitative financial data with qualitative sentiment inputs. This underscores the importance for financial analysts to craft precise and thoughtful prompts to extract optimal outputs from LLM-based tools. It also points to the growing relevance of prompt optimization as a strategic lever when deploying AI models in finance.

7.1.3 Limitations and risks

Despite these advances, the research cautions against overestimating AI capabilities. Even the most advanced AI systems, which combine sentiment analysis, technical indicators, and other data streams, only marginally outperformed traditional models in portfolio returns. Their performance largely mirrored broader market movements, suggesting that current AI cannot consistently achieve market-beating results. This finding aligns with the efficient market hypothesis: as AI adoption becomes widespread, its predictive edge may diminish due to rapid arbitrage. In addition, AI models remain vulnerable to biases in training data and to the unpredictability of real-world events, misinterpreting extreme sentiment, or failing to anticipate unprecedented market shocks. These insights emphasize that AI should serve as an aid to human decision-makers, not as a standalone tool. Firms should integrate AI with appropriate safeguards, continuously train models, and validate their performance under changing market conditions.

7.2 Asset Management

7.2.1 Portfolio construction

The findings of this thesis highlight the transformative potential of AI and sentiment analysis in finance. Integrating advanced AI models into investment operations can significantly enhance strategy optimization and portfolio management. To stay competitive, financial institutions must invest in AI infrastructure and expertise. Key applications include:

- **AI-Enhanced Portfolio Management:** Incorporating sentiment data from news and reports alongside traditional indicators allows for more dynamic and resilient portfolio strategies. AFAs that fuse price trends, fundamentals, and sentiment outperform static, rule-based systems by adapting in real time to market shifts, improving risk-adjusted returns during volatility.
- **Investment Strategy Optimization:** Traders and asset managers can integrate real-time sentiment signals from sources like social media and analyst commentary to boost forecasting accuracy. Including investor emotion indicators enables more responsive trading strategies — for instance, scaling back during sharp sentiment downturns or capitalizing on positive sentiment surges.
- **Next-Generation Robo-Advisors:** Conventional Robo-advisors struggle with sudden market shifts. Upgrading them with LLM-driven sentiment analysis can create more adaptive, resilient platforms, offering retail investors access to strategies traditionally reserved for hedge funds. Financial institutions can leverage these advanced Robo-advisors to democratize sophisticated techniques while maintaining a competitive performance edge.

The rise of AFAs signals a major evolution beyond static rule-based models. These agents interact with real-time data, adapt continuously through feedback loops, and make independent investment decisions in truly dynamic portfolio management.

At the same time, reliance on small, linear models for stock selection is increasingly inadequate. Traditional models miss the complex, nonlinear relationships between predictive signals and asset returns. In contrast, large AI and deep learning models excel at detecting these intricate patterns, a "virtue of complexity" that dramatically improves performance. Empirical results show that complex models can boost stock selection effectiveness by 50–100% in out-of-sample tests, reinforcing the transformative impact of combining AFAs with advanced analytics in modern investment management.

7.2.2 Asset management efficiency

The thesis highlights a shift from static financial advisory models to interactive, AI-driven financial services. Generative AI and sentiment analysis are positioned as key tools for democratizing access to high-quality financial advice, particularly for retail investors.

AI is increasingly used to process market data and generate trading signals tailored to individual risk profiles. Tools like *DeepSeek* have helped level the playing field, enabling smaller firms to access advanced analytics and compete with larger institutions. The research also emphasizes the effectiveness of smaller, industry-targeted language models in enhancing investment processes.

Surveys suggest that broad AI adoption in finance is still in its early stages. According to the U.S. Census Bureau, only 5% of companies currently use AI in production. However, a Federal Reserve study, adjusting for company size, estimates adoption closer to 20%, consistent with findings from several other smaller surveys. This suggests that while adoption is growing, there remains significant untapped potential for AI integration in financial services. Estimated cost savings are shown in Figure 7.1.

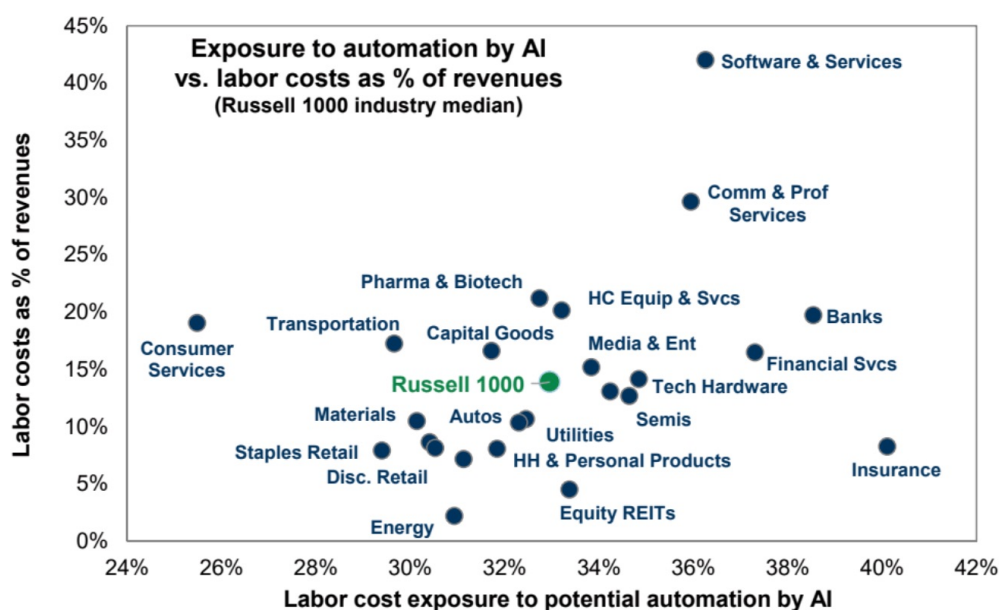


Figure 7.1: Exposure to AI automation by industry group (Hammond, 2025)

7.2.3 Portfolio risk management

Diversification remains a cornerstone of risk management, with important implications for portfolio performance. In a well-diversified portfolio, outperformance typically results from small gains across many stocks, rather than a few large wins. As shown in Figure 7.2, the central portion of a quantitative strategy's return distribution, not the outliers, drives excess returns. With sufficient diversification, strong performance is achievable even if the success rate on individual positions is only slightly above 50%. In essence, a small, consistent edge across many investments is more reliable than a concentrated, high-conviction approach.

AI enhances this dynamic by detecting shifts in market sentiment early, helping preempt downturns. Additionally, AI can validate and backtest risk models more efficiently than traditional methods, further strengthening risk management strategies.

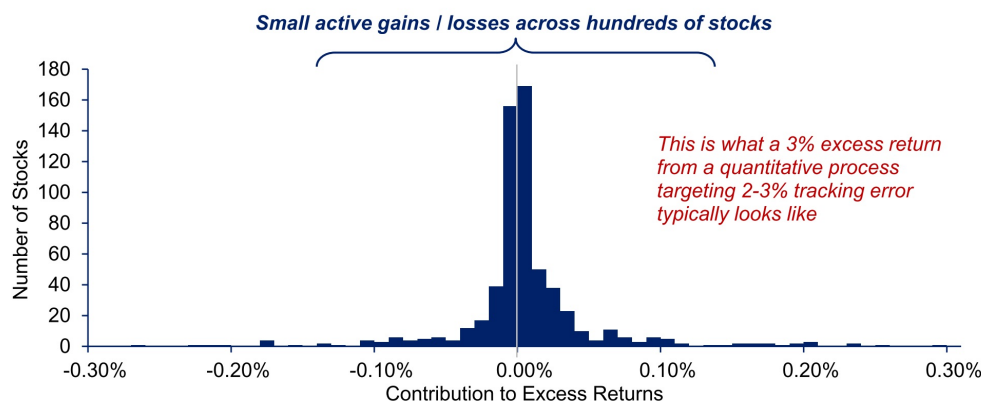


Figure 7.2: Quantitative performance is based in small gains (Fattouche et al., 2025)

7.3 Ethical and Regulatory Considerations

The advancement of AI-driven market prediction tools will impact retail and institutional investors differently. Institutional investors (i.e., hedge funds, banks, and asset managers) stand to gain the most immediate advantage, thanks to their resources to develop and deploy sophisticated AI models. By integrating sentiment-aware algorithms, institutions can execute faster, more informed trades. However, widespread adoption could create systemic risks: if many institutions react similarly to sentiment signals, herd behavior could inflate bubbles or worsen crashes. Regulators and firms will need to monitor and manage these new dynamics.

Retail investors may benefit from a trickle-down of AI technologies. Investment platforms could offer AI-generated insights and sentiment indicators, narrowing the gap between individual and institutional research capabilities. Advanced robo-advisors or personal AI financial assistants could empower retail investors with real-time, sentiment-aware strategies. However, access to top-tier AI models may remain concentrated among institutions due to their cost and complexity, risking a widening performance gap. Moreover, without proper understanding, retail users might misinterpret AI-driven advice. Investor education and transparent AI explanations will be critical to ensure responsible adoption. In short, while institutions will harness AI for a competitive edge, retail investors will need support and safeguards to fully benefit.

The deployment of AFAs also raises important regulatory and ethical considerations. Complex models like DL systems and proprietary LLMs lack transparency, making it

difficult to audit decisions or ensure regulatory compliance. Greater AI explainability will be essential, allowing firms to justify AI-driven trades to clients and regulators. Additionally, widespread reliance on sentiment analysis opens the door to manipulation (i.e., spreading false news to influence AI trading models). Safeguards and updated regulations will be needed to prevent abuse.

Bias is another concern. AI models trained on biased historical data may perpetuate or amplify systemic biases, unfairly disadvantaging certain sectors or groups. Ongoing bias audits and diverse training datasets will be necessary to promote fairness. AFAs also complicate accountability: if an AI makes flawed investment decisions as assigning responsibility becomes challenging. Financial firms must establish internal AI governance policies, while regulators may require human oversight for critical decisions. Proactive collaboration between technological and investment companies, jointly with regulatory bodies will be key to balancing innovation with financial stability, potentially through AI certification standards, mandatory stress testing, and disclosure requirements for AI-driven recommendations.

Finally, while generative AI offers transformative potential in asset management, it presents serious challenges that firms must address to ensure responsible deployment. Transparency and explainability remain central concerns, as "black box" AI models can undermine trust and regulatory compliance. Data quality and bias management are critical in building reliable models. As AI becomes embedded in investment processes, investment firms must uphold ethical standards, adapt to evolving regulations, and proactively mitigate misuse risks to maintain credibility and protect long-term market integrity.

Bibliography

- Adebiyi, A. A., Adewumi, A. O., and Ayo, C. K. (2014). Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics*, 2014:1–7.
- Akita, R., Yoshihara, A., Matsubara, T., and Uehara, K. (2016). Deep learning for stock prediction using numerical and textual information. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pages 1–6. IEEE.
- Amin, M. M., Cambria, E., and Schuller, B. W. (2023). Will Affective Computing Emerge From Foundation Models and General Artificial Intelligence? A First Evaluation of ChatGPT. *IEEE Intelligent Systems*, 38(2):15–23.
- Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models.
- Arawjo, I., Swoopes, C., Vaithilingam, P., Wattenberg, M., and Glassman, E. (2023). ChainForge: A Visual Toolkit for Prompt Engineering and LLM Hypothesis Testing.
- Arreola Hernandez, J. and Al Janabi, M. A. (2020). Forecasting of dependence, market, and investment risks of a global index portfolio. *Journal of Forecasting*, 39(3):512–532.
- Azevedo, V. and Hoegner, C. (2022). *Enhancing stock market anomalies with machine learning*. Number 0123456789. Springer US.
- Azevedo, V., Hoegner, C., and Velikov, M. (2024). The Expected Returns on Machine-Learning Strategies. *SSRN Electronic Journal*.
- Baker, M. and Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, 61(4):1645–1680.
- Baker, M. and Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives, American Economic Association*, 21(2):129–151.
- Bandyopadhyay, S. P., Brown, L. D., and Richardson, G. D. (1995). Analysts’ use of earnings forecasts in predicting stock returns: Forecast horizon effects. *International Journal of Forecasting*, 11(3):429–445.

- Bao, W., Yue, J., and Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLOS ONE*, 12(7):e0180944.
- Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343.
- Barr, M. S. (2025). Artificial Intelligence: Hypothetical Scenarios for the Future.
- Behera, J., Pasayat, A. K., Behera, H., and Kumar, P. (2023). Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. *Engineering Applications of Artificial Intelligence*, 120(December 2022):105843.
- Bernales, A., Valenzuela, M., and Zer, I. (2021). Effects of Information Overload on Financial Market Returns: How Much Is Too Much? *SSRN Electronic Journal*, 2500(1372).
- Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., and Khatri, R. K. (2022). Predicting stock market index using LSTM. *Machine Learning with Applications*, 9(May):100320.
- Bhowmik, R. and Wang, S. (2020). Stock Market Volatility and Return Analysis: A Systematic Literature Review. *Entropy*, 22(5):522.
- Bickley, S. J., Chan, H. F., and Torgler, B. (2022). Artificial intelligence in the field of economics. *Scientometrics*, 127(4):2055–2084.
- Biderman, S., Bicheno, K., and Gao, L. (2022). Datasheet for the Pile. Technical report, EleutherAI.
- Bin Ren (2024). AI in Finance: Bot, Bank & Beyond. Technical report, Citigroup.
- Black, S., Biderman, S., Hallahan, E., Anthony, Q., Gao, L., Golding, L., He, H., Leahy, C., McDonell, K., Phang, J., Pieler, M., Prashanth, U. S., Purohit, S., Reynolds, L., Tow, J., Wang, B., and Weinbach, S. (2022). GPT-NeoX-20B: An Open-Source Autoregressive Language Model. In *Proceedings of BigScience Episode #5 – Workshop on Challenges & Perspectives in Creating Large Language Models*, pages 95–136, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Boehmer, E., Fong, K., and Wu, J. J. (2021). Algorithmic Trading and Market Quality: International Evidence. *Journal of Financial and Quantitative Analysis*, 56(8):2659–2688.
- Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., and Weber, I. (2012). Web search queries can predict stock market volumes. *PloS one*, 7(7):e40014.

- Bradshaw, M. T., Brown, L. D., and Huang, K. (2013). Do sell-side analysts exhibit differential target price forecasting ability? *Review of Accounting Studies*, 18(4):930–955.
- Brenner, L. and Meyll, T. (2020). Robo-advisors: A substitute for human financial advice? *Journal of Behavioral and Experimental Finance*, 25:100275.
- Brown, E. (1999). A Non-Random Walk Down Wall Street. *Journal of Economic Surveys*, 13(4):477–478.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, 33(NeurIPS):1877–1901.
- Buchanan, B. G. (2019). Artificial Intelligence in Finance. Technical report, The Alan Turing Institute.
- Bukovina, J. (2015). Sentiment and blue-chip returns . Firm level evidence from a dynamic threshold model. Technical report, Mendel University in Brno, Faculty of Business and Economics.
- Candemir, I. and Karahan, C. C. (2024). Testing asset pricing models with individual stocks: An instrumental variables approach. *Borsa Istanbul Review*, 24(5):952–965.
- Cao, S., Jiang, W., Yang, B., and Zhang, A. (2020). How to Talk When a Machine is Listening?: Corporate Disclosure in the Age of AI. Technical report, National Bureau of Economic Research, Cambridge, MA.
- Cao, S. S., Jiang, W., Wang, J. L., and Yang, B. (2021). From Man vs. Machine to Man + Machine: The Art and Ai of Stock Analyses. *SSRN Electronic Journal*.
- CFA Institute (2019). MIFID II: ONE YEAR ON Assessing the Market for Investment Research. Technical report.
- Chatterjee, A. and Perrizo, W. (2016). Investor classification and sentiment analysis. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1177–1180. IEEE.
- Chaudhry, K. (2023). V Global Machine Learning Conference. Global Quantitative Strategy. Technical report, J.P. Morgan, Paris.

- Chaweewanchon, A. and Chaysiri, R. (2022). Markowitz Mean-Variance Portfolio Optimization with Predictive Stock Selection Using Machine Learning. *International Journal of Financial Studies*, 10(3):64.
- Chen, B., Zhang, Z., Langrené, N., and Zhu, S. (2023a). Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review.
- Chen, K., Zhou, Y., and Dai, F. (2015). A LSTM-based method for stock returns prediction: A case study of China stock market. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 2823–2824. IEEE.
- Chen, M., DeHaven, M., Kitschelt, I., Lee, S. J., and Sicilian, M. J. (2023b). Identifying Financial Crises Using Machine Learning on Textual Data. *Journal of Risk and Financial Management*, 16(3):161.
- Chen, W., Su, Y., Zuo, J., Yang, C., Yuan, C., Chan, C.-M., Yu, H., Lu, Y., Hung, Y.-H., Qian, C., Qin, Y., Cong, X., Xie, R., Liu, Z., Sun, M., and Zhou, J. (2023c). Agent-Verse: Facilitating Multi-Agent Collaboration and Exploring Emergent Behaviors. In *ICLR 2024*.
- Cheng, Y., Zhang, C., Zhang, Z., Meng, X., Hong, S., Li, W., Wang, Z., Wang, Z., Yin, F., Zhao, J., and He, X. (2024). Exploring Large Language Model based Intelligent Agents: Definitions, Methods, and Prospects.
- Cho, P., Park, J. H., and Song, J. W. (2021). Equity Research Report-Driven Investment Strategy in Korea Using Binary Classification on Stock Price Direction. *IEEE Access*, 9:46364–46373.
- Chopra, R. and Sharma, G. D. (2021). Application of Artificial Intelligence in Stock Market Forecasting: A Critique, Review and Research Agenda. *Journal of Risk and Financial Management*, 14(11):526.
- Chowdhury, E. K., Dhar, B. K., and Stasi, A. (2022). Volatility of the US stock market and business strategy during COVID-19. *Business Strategy & Development*, 5(4):350–360.
- Clarke, R. G., de Silva, H., and Thorley, S. (2006). Minimum-Variance Portfolios in the U.S. Equity Market. *The Journal of Portfolio Management*, 33(1):10–24.
- Clavié, B., Ciceu, A., Naylor, F., Soulié, G., and Brightwell, T. (2023). Large Language Models in the Workplace: A Case Study on Prompt Engineering for Job Type Classification.
- Cook, V. (2024). Transformation AI: From evolution to revolution? Technical Report September, Bank of America Institute.

- Cook, V. (2025). The new wave: Agentic AI. Technical Report January, Bank of America Institute.
- Corbet, S., Dowling, M., and Cummins, M. (2015). Analyst recommendations and volatility in a rising, falling, and crisis equity market. *Finance Research Letters*, 15:187–194.
- Dacorogna, M. M., Gencay, R., Müller, U. A., and Pictet, O. V. (1999). Effective Return, Risk Aversion and Drawdowns. *SSRN Electronic Journal*, 289(519):229–248.
- Dahal, K. R., Pokhrel, N. R., Gaire, S., Mahatara, S., Joshi, R. P., Gupta, A., Banjade, H. R., and Joshi, J. (2023). A comparative study on effect of news sentiment on stock price prediction with deep learning architecture. *PLoS ONE*, 18(4 April):1–19.
- Dangl, T. and Salbrechter, S. (2021). News Sentiment and Equity Returns. *SSRN Electronic Journal*, pages 1–46.
- Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor Psychology and Security Market Under- and Overreactions. *The Journal of Finance*, 53(6):1839–1885.
- Das, A., Kong, W., Sen, R., and Zhou, Y. (2023). A decoder-only foundation model for time-series forecasting.
- Dash, R. and Dash, P. K. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques. *The Journal of Finance and Data Science*, 2(1):42–57.
- Dell’Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Kraymer, L., Candelon, F., and Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm):4171–4186.
- Dingli, A. and Fournier, K. S. (2017). Financial time series forecasting - a deep learning approach. *International Journal of Machine Learning and Computing*, 7(5):118–122.
- Dreman, D., Johnson, S., MacGregor, D., and Slovic, P. (2001). A Report on the March 2001 Investor Sentiment Survey. *Journal of Psychology and Financial Markets*, 2(3):126–134.
- Duz Tan, S. and Tas, O. (2021). Social Media Sentiment in International Stock Returns and Trading Activity. *Journal of Behavioral Finance*, 22(2):221–234.

- Ehrmann, M. (2004). Taking Stock: Monetary Policy Transmission to Equity Markets.
- El Alaoui, M., Bouri, E., and Azoury, N. (2020). The Determinants of the U.S. Consumer Sentiment: Linear and Nonlinear Models. *International Journal of Financial Studies*, 8(3):38.
- Eva Dezsai (2016). Can deep machine learning outsmart the market? A comparison between econometric modelling and LSTM. *Romanian Economic and Business Review*, December:54–73.
- Evdokimov, A. (2024). Do Prompt Structures Improve Output Quality? Testing Zero-Shot Prompts with GPT-4, Claude 3 and Gemini 1.5.
- Fallahgoul, H. A. and Lin, X. (2021). Who Influences Whom? Behavior Contagion among Investors. *SSRN Electronic Journal*.
- Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2):383.
- Fatouros, G., Soldatos, J., Kouroumalis, K., Makridakis, G., and Kyriazis, D. (2023). Transforming Sentiment Analysis in the Financial Domain with ChatGPT. *Machine Learning with Applications*, 14(October):100508.
- Fattouche, C., Maloney, T., and Brixton, A. (2025). A New Paradigm in Active Equity: Navigating Concentrated Markets and Harnessing New Technologies. Technical report, AQR.
- Frankel, R., Jennings, J., and Lee, J. (2022). Disclosure Sentiment: Machine Learning vs. Dictionary Methods. *Management Science*, 68(7):5514–5532.
- Frijns, B. and Garel, A. (2021). The effect of cultural distance between an analyst and a CEO on analysts’ earnings forecast performance. *Economics Letters*, 205:109957.
- Gangopadhyay, S. and Majumder, P. (2023). Text representation for direction prediction of share market. *Expert Systems with Applications*, 211(August 2022):118472.
- Gao, F. and Han, L. (2012). Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications*, 51(1):259–277.
- Gao, R., Cui, S., Xiao, H., Fan, W., Zhang, H., and Wang, Y. (2022). Integrating the sentiments of multiple news providers for stock market index movement prediction: A deep learning approach based on evidential reasoning rule. *Information Sciences*, 615:529–556.
- Gao, Y., Wang, R., and Zhou, E. (2021). Stock Prediction Based on Optimized LSTM and GRU Models. *Scientific Programming*, 2021:1–8.

- Garlick, R., Spielman, A., Wilkie, M., Krause, H., Moffat, M., Gibson, C., and Brown, M. (2024). The Rise of AI Robots. Technical Report December, Citigroup.
- Garza, A. and Mergenthaler-Canseco, M. (2023). TimeGPT-1.
- Ghose, R. (2025). Agentic AI. Technical Report January, Citigroup.
- Goktas, F. and Duran, A. (2019). A new possibilistic mean – Variance model based on the principal components analysis: An application on the Turkish holding stocks. *Journal of Multiple-Valued Logic and Soft Computing*, 32(5-6):455–476.
- Gordon, M. V. and Lunsford, K. G. (2024). The effects of the Federal Reserve Chair’s testimony on interest rates and stock prices. *Economics Letters*, 235(November 2023):111537.
- Grant, A., Jarnećić, E., and Su, M. (2015). Asymmetric effects of sell-side analyst optimism and broker market share by clientele. *Journal of Financial Markets*, 24:49–65.
- Grealish, A. and Kolm, P. N. (2021). Robo-Advisory: From Investing Principles and Algorithms to Future Developments. *SSRN Electronic Journal*, pages 1–29.
- Gruver, N., Finzi, M., Qiu, S., and Wilson, A. G. (2023). Large Language Models Are Zero-Shot Time Series Forecasters. In *37th Conference on Neural Information Processing Systems (NeurIPS 2023)*, number NeurIPS, pages 1–29.
- Gu, J., Shukla, S., Ye, J., Uddin, A., and Wang, G. (2023). Deep learning model with sentiment score and weekend effect in stock price prediction. *SN Business & Economics*, 3(7):119.
- Guerard, J. B., Markowitz, H., and Xu, G. (2015). Earnings forecasting in a global stock selection model and efficient portfolio construction and management. *International Journal of Forecasting*, 31(2):550–560.
- Guo, X., Gu, C., Zhang, C., and Li, S. (2024). Institutional herding and investor sentiment. *Journal of Financial Markets*, 68(January):100891.
- Guo, Z., Jiang, G., Zhang, Z., Li, P., Wang, Z., and Wang, Y. (2023). Shai: A large language model for asset management.
- Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., and Smith, N. A. (2020). Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Stroudsburg, PA, USA. Association for Computational Linguistics.

- Hammond, R. (2025). AI and US Equities: Refreshing our list of AI Phase 3 stocks and identifying early adopters in AI Phase 4. Technical report, Goldman Sachs.
- Hatzius, J., Briggs, J., Dong, S., and Peters, M. (2024). AI Adoption Tracker. Technical Report December 2024, Goldman Sachs.
- Heaton, J. B., Polson, N. G., and Witte, J. H. (2017). Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1):3–12.
- Hendershott, T. and Riordan, R. (2013). Algorithmic Trading and the Market for Liquidity. *Journal of Financial and Quantitative Analysis*, 48(4):1001–1024.
- Henrique, B. M., Sobreiro, V. A., and Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 124:226–251.
- Heston, S. L. and Sinha, N. R. (2016). News versus Sentiment: Predicting Stock Returns from News Stories. *Finance and Economics Discussion Series*, 2016(048):1–35.
- Hira, Z. M. and Gillies, D. F. (2015). A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. *Advances in Bioinformatics*, 2015(1):1–13.
- Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., and Soman, K. P. (2018). NSE Stock Market Prediction Using Deep-Learning Models. In *Procedia Computer Science*, volume 132.
- Holtfort, T., Horsch, A., and Schwarz, J. (2022). Robo-advisors as part of industry 4.0 in financial markets: Evolutionary development, methods, and first performance insights. *Risk Governance and Control: Financial Markets and Institutions*, 12(2):38–46.
- Hong, S., Zhuge, M., Chen, J., Zheng, X., Cheng, Y., Zhang, C., Wang, J., Wang, Z., Yau, S. K. S., Lin, Z., Zhou, L., Ran, C., Xiao, L., Wu, C., and Schmidhuber, J. (2023). MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework.
- Hsu, M.-W., Lessmann, S., Sung, M.-C., Ma, T., and Johnson, J. E. (2016). Bridging the divide in financial market forecasting: machine learners vs. financial economists. *Expert Systems with Applications*, 61(May):215–234.
- Hu, B., Zhao, C., Zhang, P., Zhou, Z., Yang, Y., Xu, Z., and Liu, B. (2023). Enabling Intelligent Interactions between an Agent and an LLM: A Reinforcement Learning Approach.
- Huang, S., Wu, X., Wang, D., and Tang, A. (2019). DSANet: Dual self-attention network for multivariate time series forecasting. *International Conference on Information and Knowledge Management, Proceedings*, pages 2129–2132.

- Huang, Y., Zhou, C., Cui, K., and Lu, X. (2024). A multi-agent reinforcement learning framework for optimizing financial trading strategies based on TimesNet. *Expert Systems with Applications*, 237(July 2023):121502.
- Huang, Z. and Tanaka, F. (2022). MSPM: A modularized and scalable multi-agent reinforcement learning-based system for financial portfolio management. *PLOS ONE*, 17(2):e0263689.
- Jang, J. and Seong, N. (2023). Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory. *Expert Systems with Applications*, 218(November 2022):119556.
- Jeff Cao, Pete Hecht, B. K. N. M. (2024). Can Machines Time Markets? Technical Report 2, AQR.
- Jiang, W. (2021). Applications of deep learning in stock market prediction: Recent progress. *Expert Systems with Applications*, 184(March 2020):115537.
- Jiang, X. and Li, S. (2017). BAS: Beetle Antennae Search Algorithm for Optimization Problems. *International Journal of Robotics and Control*, 1(1):1.
- J.P. Morgan (2025). JPMorganChase: 225 years of history.
- Kaczmarek, K. (2023). The generative AI tipping point. Technical report, Oliver Wyman.
- Kaplanski, G., Levy, H., Veld, C., and Veld-Merkoulova, Y. (2015). Do Happy People Make Optimistic Investors? *Journal of Financial and Quantitative Analysis*, 50(1-2):145–168.
- Karathanasopoulos, A., Mitra, S., Skindilias, K., and Lo, C. C. (2017). Modelling and Trading the English and German Stock Markets with Novelty Optimization Techniques. *Journal of Forecasting*, 36(8):974–988.
- Karmaziene, E. (2023). The greater the volume, the greater the analyst. *Finance Research Letters*, 51(August 2022):103377.
- Keohane, D. (2018). French watchdog calls for rethink of research rules.
- Kheiri, K. and Karimi, H. (2023). SentimentGPT: Exploiting GPT for Advanced Sentiment Analysis and its Departure from Current Machine Learning.
- Kim, A. G., Muhn, M., Nikolaev, V. V., and Kim, A. G. (2023a). Bloated Disclosures: Can ChatGPT Help Investors Process Information? *SSRN Electronic Journal*, (23).
- Kim, S., Lee, W.-J., Park, S., and Sunwoo, H.-Y. (2022). Busy analysts in uncertain times. *Finance Research Letters*, 46(PB):102488.

- Kim, S. S. S., Kim, S. S. S., Kim, Y., Park, J., Kim, S. S. S., Kim, M., Sung, C. H., Hong, J., and Lee, Y. (2023b). LLMs Analyzing the Analysts: Do BERT and GPT Extract More Value from Financial Analyst Reports? In *4th ACM International Conference on AI in Finance*, pages 383–391, New York, NY, USA. ACM.
- Kirtac, K. and Germano, G. (2024). Sentiment trading with large language models. *Finance Research Letters*, 62(PB):105227.
- Ko, H. and Lee, J. (2023). Can ChatGPT Improve Investment Decision? From a portfolio manager perspective.
- Kobiela, D., Krefta, D., Król, W., and Weichbroth, P. (2022). ARIMA vs LSTM on NASDAQ stock exchange data. *Procedia Computer Science*, 207(Kes):3836–3845.
- Kolonin, A., Glushchenko, A., Fokin, A., Mari, M., Casiraghi, M., and Vishwas, M. (2023). Adaptive Predictive Portfolio Management Agent. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 13921 LNCS, pages 187–196.
- Krause, D. (2023a). Large Language Models and Generative AI in Finance: An Analysis of ChatGPT, Bard, and Bing AI. *SSRN Electronic Journal*, pages 1–31.
- Krause, D. (2023b). Proper Generative AI Prompting for Financial Analysis. *SSRN Electronic Journal*, pages 1–28.
- Krollner, B., Vanstone, B., and Finnie, G. (2010). Financial time series forecasting with machine learning techniques: A survey. In *Proceedings of the 18th European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning, ESANN 2010*, pages 25–30.
- Kumar, R. and Sharma, V. (2023). The AI Revolution. Technical Report November, COATUE.
- Laurençon, H., Saulnier, L., Wang, T., Akiki, C., Moral, V., Scao, T. L., Werra, L. V., Mou, C., Ponferrada, E. G., Nguyen, H., Laurençon, H., Saulnier, L., Wang, T., Akiki, C., and Villanova, A. (2022). The BigScience ROOTS Corpus: A 1 . 6TB Composite Multilingual Dataset. In *Advances in Neural Information Processing Systems*, number 35, pages 31809–31826.
- Le, T. D. and Trinh, T. (2022). Distracted analysts and earnings management. *Finance Research Letters*, 49(June):103038.
- Lee, S. I. and Yoo, S. J. (2020). Threshold-based portfolio: the role of the threshold and its applications. *The Journal of Supercomputing*, 76(10):8040–8057.
- Lemmon, M. and Portniaguina, E. (2006). Consumer confidence and asset prices: some empirical evidence. *Review of Financial Studies*, 19(4):1499–1529.

- Li, E. X., Tu, Z., and Zhou, D. (2023a). The Promise and Peril of Generative AI: Evidence from ChatGPT as Sell-Side Analysts. *SSRN Electronic Journal*, (June):1–31.
- Li, F. (2006). Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports? *SSRN Electronic Journal*.
- Li, H., Shen, Y., and Zhu, Y. (2018). Stock Price Prediction Using Attention-based Multi-Input LSTM. In *Proceedings of Machine Learning Research*, volume 95, pages 454–469.
- Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., and Chen, Y. (2014). The effect of news and public mood on stock movements. *Information Sciences*, 278:826–840.
- Li, T. (2022). Analyst’s stock views and revision actions. *Finance Research Letters*, 44(April 2021):102033.
- Li, X., Wu, P., and Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing & Management*, 57(5):102212.
- Li, X., Zhao, R., Chia, Y. K., Ding, B., Joty, S., Poria, S., and Bing, L. (2024). Chain-of-Knowledge: Grounding Large Language Models via Dynamic Knowledge Adapting over Heterogeneous Sources.
- Li, Y., Lv, S., Liu, X., and Zhang, Q. (2022). Incorporating Transformers and Attention Networks for Stock Movement Prediction. *Complexity*, 2022.
- Li, Y., Wang, S., Ding, H., and Chen, H. (2023b). Large Language Models in Finance: A Survey. In *4th ACM International Conference on AI in Finance*, number in, pages 374–382, New York, NY, USA. ACM.
- Li, Y., Yu, Y., Li, H., Chen, Z., and Khashanah, K. (2023c). TradingGPT: Multi-Agent System with Layered Memory and Distinct Characters for Enhanced Financial Trading Performance.
- Lilian Weng (2023a). <https://lilianweng.github.io/posts/2023-06-23-agent/>.
- Lilian Weng (2023b). <https://www.linkedin.com/in/lilianweng>.
- Lin, B. Y., Fu, Y., Yang, K., Brahman, F., Huang, S., Bhagavatula, C., Ammanabrolu, P., Choi, Y., and Ren, X. (2023). SwiftSage: A Generative Agent with Fast and Slow Thinking for Complex Interactive Tasks. (NeurIPS):1–18.
- Lin, W. and Taamouti, A. (2024). Portfolio selection under non-gaussianity and systemic risk: A machine learning based forecasting approach. *International Journal of Forecasting*, 40(3):1179–1188.

- Lin, Y., Liu, S., Yang, H., Wu, H., and Jiang, B. (2021). Improving stock trading decisions based on pattern recognition using machine learning technology. *PLOS ONE*, 16(8):e0255558.
- Ling, C., Zhao, X., Lu, J., Deng, C., Zheng, C., Wang, J., Chowdhury, T., Li, Y., Cui, H., Zhang, X., Zhao, T., Panalkar, A., Cheng, W., Wang, H., Liu, Y., Chen, Z., Chen, H., White, C., Gu, Q., Pei, J., and Zhao, L. (2023). Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey.
- Ling, D. C., Naranjo, A., and Scheick, B. (2014). Investor Sentiment, Limits to Arbitrage and Private Market Returns. *Real Estate Economics*, 42(3):531–577.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G. (2023a). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, 55(9):1–35.
- Liu, S., Zhang, C., and Ma, J. (2017). CNN-LSTM Neural Network Model for Quantitative Strategy Analysis in Stock Markets. In *Neural Information Processing. ICONIP 2017. Lecture Notes in Computer Science*, pages 198–206. Springer.
- Liu, X.-y., Wang, G., Yang, H., and Zha, D. (2023b). FinGPT: Democratizing Internet-scale Data for Financial Large Language Models.
- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., and Ge, B. (2023c). Summary of ChatGPT-Related research and perspective towards the future of large language models. *Meta-Radiology*, 1(2):100017.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach.
- Liu, Z., Zhang, Y., Li, P., Liu, Y., and Yang, D. (2023d). Dynamic LLM-Agent Network: An LLM-agent Collaboration Framework with Agent Team Optimization.
- Lo, H.-C. (2017). Do Firm Size Influence Financial Analyst Research Reports and Subsequent Stock Performance. *Accounting and Finance Research*, 6(4):181.
- Long, J. B. D., Shleifer, A., Summers, L. H., Waldmann, R. J., De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4):703.
- Long, W., Lu, Z., and Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems*, 164:163–173.

- Löf, H., Sahamkhadam, M., and Stephan, A. (2023). Incorporating ESG into Optimal Stock Portfolios for the Global Timber and Forestry Industry. *Journal of Forest Economics*, 38(2):133–157.
- Lopez-lira, A. and Tang, Y. (2023). Can ChatGPT Forecast Stock Price Movements ? Return Predictability and Large Language Models. *SSRN Electronic Journal*.
- Loughgran, T., McDonald, B., and Loughran; T.; & McDonald; B (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, 66(1):35–65.
- Loughran, T. and McDonald, B. (2020). Textual Analysis in Finance. *Annual Review of Financial Economics*, 12(1):357–375.
- Lu, T., Shiu, Y., and Liu, T. (2012). Profitable candlestick trading strategies—The evidence from a new perspective. *Review of Financial Economics*, 21(2):63–68.
- Luo, Y., Tang, Y., Shen, C., Zhou, Z., and Dong, B. (2023). Prompt Engineering Through the Lens of Optimal Control. *Journal of Machine Learning*, 2(4):241–258.
- Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, 17(1):59–82.
- Mao, Y., Wei, W., Wang, B., and Liu, B. (2012). Correlating S&P 500 stocks with Twitter data. In *Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research - HotSocial '12*, pages 69–72, New York, New York, USA. ACM Press.
- Marrs, S. (2024). Predictions 2025: GenAI As A Growth Driver Will Put B2B Executives To The Test.
- McMillan, D. G. (2016). Which Variables Predict and Forecast Stock Market Returns? *SSRN Electronic Journal*, 44(June):0–28.
- Megaw, N. (2023). Investors use AI to glean signals behind executives ’ soothing words. *Financial Times*.
- Mirete-Ferrer, P. M., Garcia-Garcia, A., Baixauli-Soler, J. S., and Prats, M. A. (2022). A Review on Machine Learning for Asset Management. *Risks*, 10(4):84.
- Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., and Trajanov, D. (2020). Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access*, 8:131662–131682.
- Mohan, S., Mullanpudi, S., Sammeta, S., Vijayvergia, P., and Anastasiu, D. C. (2019). Stock Price Prediction Using News Sentiment Analysis. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 205–208. IEEE.

- Morris, M. R., Sohl-dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., and Legg, S. (2023). Levels of AGI: Operationalizing Progress on the Path to AGI.
- Muktadir, G. M. (2023). A Brief History of Prompt: Leveraging Language Models (Through Advanced Prompting).
- Myšková, R. and Hájek, P. (2020). Mining risk-related sentiment in corporate annual reports and its effect on financial performance. *Technological and Economic Development of Economy*, 26(6):1422–1443.
- Nazareth, N. and Ramana Reddy, Y. V. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications*, 219(January 2022):119640.
- Neslihanoglu, S., Sogiakas, V., McColl, J. H., and Lee, D. (2017). Nonlinearities in the CAPM: Evidence from Developed and Emerging Markets. *Journal of Forecasting*, 36(8):867–897.
- Nguyen, T. P. L., Chew, L. W., Muthaiyah, S., Teh, B. H., and Ong, T. S. (2023). Factors influencing acceptance of Robo-Advisors for wealth management in Malaysia. *Cogent Engineering*, 10(1).
- Nison, S. (1991). *Japanese Candlestick Charting Techniques: a Contemporary Guide to the Ancient Investment Techniques of the Far East*. New York Institute of Finance, New York.
- Nyakurukwa, K. and Seetharam, Y. (2023). Can textual sentiment partially explain differences in the prices of dual-listed stocks? *Finance Research Letters*, 58(PC):104529.
- Olof, L. (2019). *Sentiment Analysis of Equity Analyst Research Reports using Convolutional Neural Networks*. PhD thesis, Uppsala University.
- Pasch, S. and Ehnes, D. (2022). StonkBERT: Can Language Models Predict Medium-Run Stock Price Movements?
- Patil, D., Patil, S., Patil, S., and Arora, S. (2022). Financial Forecasting of Stock Market Using Sentiment Analysis and Data Analytics. In *Lecture Notes in Networks and Systems*, volume 334, pages 423–430.
- Pelster, M. and Val, J. (2024). Can ChatGPT assist in picking stocks? *Finance Research Letters*, 59(November 2023):104786.
- Peng, B., Galley, M., He, P., Cheng, H., Xie, Y., Hu, Y., Huang, Q., Liden, L., Yu, Z., Chen, W., and Gao, J. (2023). Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback.
- Poggio, T., Mhaskar, H., Rosasco, L., Miranda, B., and Liao, Q. (2017). Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review. *International Journal of Automation and Computing*, 14(5):503–519.

- Pursiainen, V. (2018). Cultural Biases in Equity Analysis. *SSRN Electronic Journal*.
- Qian, C., Cong, X., Liu, W., Yang, C., Chen, W., Su, Y., Dang, Y., Li, J., Xu, J., Li, D., Liu, Z., and Sun, M. (2023). Communicative Agents for Software Development.
- Qiu, J., Wang, B., and Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PLoS ONE*, 15(1):1–15.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. *OpenAI blog*.
- Rae, J. W., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Song, F., Aslanides, J., Henderson, S., Ring, R., Young, S., Rutherford, E., Hennigan, T., Menick, J., Cassirer, A., Powell, R., Driessche, G. v. d., Hendricks, L. A., Rauh, M., Huang, P.-S., Glaese, A., Welbl, J., Dathathri, S., Huang, S., Uesato, J., Mellor, J., Higgins, I., Creswell, A., McAleese, N., Wu, A., Elsen, E., Jayakumar, S., Buchatskaya, E., Budden, D., Sutherland, E., Simonyan, K., Paganini, M., Sifre, L., Martens, L., Li, X. L., Kuncoro, A., Nematzadeh, A., Gribovskaya, E., Donato, D., Lazaridou, A., Mensch, A., Lespiau, J.-B., Tsimpoukelli, M., Grigorev, N., Fritz, D., Sottiaux, T., Pajarskas, M., Pohlen, T., Gong, Z., Toyama, D., D’Autume, C. d. M., Li, Y., Terzi, T., Mikulik, V., Babuschkin, I., Clark, A., Casas, D. d. L., Guy, A., Jones, C., Bradbury, J., Johnson, M., Hechtman, B., Weidinger, L., Gabriel, I., Isaac, W., Lockhart, E., Osindero, S., Rimell, L., Dyer, C., Vinyals, O., Ayoub, K., Stanway, J., Bennett, L., Hassabis, D., Kavukcuoglu, K., and Irving, G. (2022). Scaling Language Models: Methods, Analysis & Insights from Training Gopher.
- Rawte, V., Sheth, A., and Das, A. (2023). A Survey of Hallucination in Large Foundation Models.
- Reid, H. (2019). www.reuters.com/article/business/pressure-on-small-brokers-grows-a-year-after-new-eu-rules.
- Risk Control Limited (2020). The impact of MiFID II on SME and fixed income investment research Final report. Technical Report April, European Commission.
- Rose, D., Himakunthala, V., Ouyang, A., He, R., Mei, A., Lu, Y., Saxon, M., Sonar, C., Mirza, D., and Wang, W. Y. (2023). Visual Chain of Thought: Bridging Logical Gaps with Multimodal Infillings.
- Rybinski, K. (2020a). Should asset managers pay for economic research? A machine learning evaluation. *The Journal of Finance and Data Science*, 6:31–48.
- Rybinski, K. (2020b). The forecasting power of the multi-language narrative of sell-side research: A machine learning evaluation. *Finance Research Letters*, 34(August 2019):101261.

- Saetia, K. and Yokrattanasak, J. (2022). Stock Movement Prediction Using Machine Learning Based on Technical Indicators and Google Trend Searches in Thailand. *International Journal of Financial Studies*, 11(1):5.
- Sahani, T. (2024). Decoding Market Emotions: The Synergy of Sentiment Analysis and AI in Stock Market Predictions. *Journal of Next-Generation Research* 5.0, 1(1):1–11.
- Samarawickrama, A. and Fernando, T. (2017). A recurrent neural network approach in predicting daily stock prices an application to the Sri Lankan stock market. In *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, pages 1–6. IEEE.
- Schlaubitz, A. (2021). *Natural Language Processing in finance: analysis of sentiment and complexity of news and earnings reports of swiss SMEs and their relevance for stock returns*. PhD thesis, Zurich University of Applied Sciences.
- Schumpeter, J. A. and Keynes, J. M. (1936). The General Theory of Employment, Interest and Money. *Journal of the American Statistical Association*, 31(196):791.
- Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., and Soman, K. P. (2017). Stock price prediction using LSTM, RNN and CNN-sliding window model. *2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017*, 2017-Janua(September):1643–1647.
- SemEval: The 18th International Workshop on Semantic Evaluation (2024). <https://semeval.github.io/SemEval2024>.
- Sen, J., Dutta, A., and Mehtab, S. (2021). Stock Portfolio Optimization Using a Deep Learning LSTM Model. In *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*, pages 263–271. IEEE.
- Sezer, O. B., Gudelek, M. U., and Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing*, 90(December):106181.
- Shah, D., Campbell, W., and Zulkernine, F. H. (2019). A Comparative Study of LSTM and DNN for Stock Market Forecasting. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, (December):4148–4155.
- Shen, Y., Liu, C., Sun, X., and Guo, K. (2023). Investor sentiment and the Chinese new energy stock market: A risk–return perspective. *International Review of Economics & Finance*, 84(November 2022):395–408.
- Sheridan, E. (2025). Generative AI Part X: Examining the Landscape in the Face of Open-Sourced Model Performance. Technical report, Goldman Sachs, New York.

- Shih, S. Y., Sun, F. K., and Lee, H. y. (2019). Temporal pattern attention for multivariate time series forecasting. *Machine Learning*, 108(8-9):1421–1441.
- Shiller, R. J. (2003). From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perspectives*, 17(1):83–104.
- Shöhnke M Bartram (2020). Artificial Intelligence in Asset Management. Technical report, CFA Institute Research Foundation.
- Sidogi, T., Mongwe, W. T., Mbuva, R., and Marwala, T. (2022). Fusing Sell-Side Analyst Bidirectional Forecasts Using Machine Learning. *IEEE Access*, 10(July):76966–76974.
- Sirignano, J., Sadhwani, A., and Giesecke, K. (2018). Deep Learning for Mortgage Risk. *SSRN Electronic Journal*, pages 1–83.
- Souza, T. T. P., Kolchyna, O., Treleaven, P. C., Aste, T., Treleaven, P. C., and Aste, T. (2015). Twitter Sentiment Analysis Applied to Finance: A Case Study in the Retail Industry. *arXiv*, (July):19.
- Stanley, M., Stanley, E., Ovrum, M., Duverce, B., Woodring, E. W., Pauker, A. B., Grenu, R., Weiss, K., Nowak, B., Thun, T. J., Neuman, M. C., Keyser, K. A., Byrd, S. C., Limited+, A., Kim, S., and Castagno, T. (2024). Thematics: AI: Adopters Are (Still) Adopting. Technical report, Morgan Stanley.
- Sul, H. K., Dennis, A. R., and Yuan, L. I. (2017). Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns. *Decision Sciences*, 48(3):454–488.
- Sun, C., Qiu, X., Xu, Y., and Huang, X. (2019). How to Fine-Tune BERT for Text Classification? In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11856 LNAI, pages 194–206.
- Suzuki, M., Sakaji, H., Izumi, K., and Ishikawa, Y. (2022). Forecasting Stock Price Trends by Analyzing Economic Reports With Analyst Profiles. *Frontiers in Artificial Intelligence*, 5(June).
- Swathi, T., Kasiviswanath, N., and Rao, A. A. (2022). An optimal deep learning-based LSTM for stock price prediction using twitter sentiment analysis. *Applied Intelligence*, 52(12):13675–13688.
- Talebirad, Y. and Nadiri, A. (2023). Multi-Agent Collaboration: Harnessing the Power of Intelligent LLM Agents.
- Tang, B. and Matteson, D. S. (2021). Probabilistic Transformer for Time Series Analysis. *Advances in Neural Information Processing Systems*, 28(NeurIPS):23592–23608.

- Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62(3):1139–1168.
- Thas Thaker, H. M., Mohamad, A., Mustaffa Kamil, N. K., and Duasa, J. (2018). How useful are the equity analysts’ report? Evidence from Malaysia. *Reports on Economics and Finance*, 4(4):221–246.
- Thorp, W. A. (2004). www.aaii.com/journal/article/feature-investor-sentiment-as-a-contrarian-indicator.
- Tiberius, V., Gojowy, R., and Dabić, M. (2022). Forecasting the future of robo advisory: A three-stage Delphi study on economic, technological, and societal implications. *Technological Forecasting and Social Change*, 182(October 2021):121824.
- Trading System (2024). <https://therobusttrader.com/what-percentage-of-trading-is-algorithmic/>.
- Tumminello, M., Aste, T., Di Matteo, T., and Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*, 102(30):10421–10426.
- Useche, A. J., Martínez-Ferrero, J., and Alayón-Gonzales, J. L. (2024). Socially responsible portfolios, environmental, social, corporate governance (ESG) efficient frontiers and psychic dividends. *Corporate Social Responsibility and Environmental Management*, 31(2):1323–1339.
- Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., and Sandoval-Almazán, R. (2022). Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study of the H1N1 and the COVID-19 Periods. *Cognitive Computation*, 14(1):372–387.
- van Binsbergen, J. H., Han, X., and Lopez-Lira, A. (2021). Textual Analysis of Short-seller Research Reports, Stock Prices, and Real Investment. *SSRN Electronic Journal*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention Is All You Need. *IEEE Industry Applications Magazine*, 8(1):8–15.
- Velásquez-Henao, J. D., Franco-Cardona, C. J., and Cadavid-Higuaita, L. (2023). Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering. *DYNA*, 90(230):9–17.
- Wang, C., Chen, Y., Zhang, S., and Zhang, Q. (2022). Stock market index prediction using deep Transformer model. *Expert Systems with Applications*, 208(July):118128.
- Wang, K., Lu, Y., Santacroce, M., Gong, Y., Zhang, C., and Shen, Y. (2023a). Adapting LLM Agents Through Communication.

- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., and Wen, J.-R. (2023b). A Survey on Large Language Model based Autonomous Agents.
- Wang, Y.-H., Keswani, A., and Taylor, S. J. (2006). The relationships between sentiment, returns and volatility. *International Journal of Forecasting*, 22(1):109–123.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. (NeurIPS):1–43.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., and Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT.
- Williams, L. (2025). Enterprise AI predictions 2025: The year GenAI creates tangible value?
- Wisniewski, T. P. and Yekini, L. S. (2015). Stock market returns and the content of annual report narratives. *Accounting Forum*, 39(4):281–294.
- Wu, S., Irsoy, O., Lu, S., Dabrowski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., and Mann, G. (2023). BloombergGPT: A Large Language Model for Finance.
- Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., Yin, Z., Dou, S., Weng, R., Cheng, W., Zhang, Q., Qin, W., Zheng, Y., Qiu, X., Huang, X., and Gui, T. (2023). The Rise and Potential of Large Language Model Based Agents: A Survey.
- Xie, Q., Han, W., Zhang, X., Lai, Y., Peng, M., Lopez-Lira, A., and Huang, J. (2023). PIXIU: A Large Language Model, Instruction Data and Evaluation Benchmark for Finance.
- Yang, G., He, P., and Liu, X. (2023a). FundRecLLM: Fund Recommendation Based on Financial News and Research Analyst Report. *Frontiers in Artificial Intelligence and Applications*, 378:515–523.
- Yang, H., Liu, X.-Y., and Dan Wang, C. (2023b). FinGPT: Open-Source Financial Large Language Models. *SSRN Electronic Journal*.
- Yang, H., Zhang, B., Wang, N., Guo, C., Zhang, X., Lin, L., Wang, J., Zhou, T., Guan, M., Zhang, R., and Wang, C. D. (2024). FinRobot: An Open-Source AI Agent Platform for Financial Applications using Large Language Models.

- Yang, N., Fernandez-Perez, A., and Indriawan, I. (2022). The Price Impact of Tweets: A High-Frequency Study. *SSRN Electronic Journal*.
- Yang, Y., UY, M. C. S., and Huang, A. (2020). FinBERT: A Pretrained Language Model for Financial Communications.
- Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., and Narasimhan, K. (2023). Tree of Thoughts: Deliberate Problem Solving with Large Language Models. In *37th Conference on Neural Information Processing Systems (NeurIPS 2023)*, number 1, pages 1–14.
- Yu, Y., Li, H., Chen, Z., Jiang, Y., Li, Y., Zhang, D., Liu, R., Suchow, J. W., and Khashanah, K. (2024). FinMem: A Performance-Enhanced LLM Trading Agent with Layered Memory and Character Design. In *Proceedings of the AAAI Symposium Series*, volume 3, pages 595–597.
- Yue, T., Au, D., Au, C. C., and Iu, K. Y. (2023). Democratizing Financial Knowledge with ChatGPT by OpenAI: Unleashing the Power of Technology. *SSRN Electronic Journal*, pages 1–26.
- Zeng, A., Chen, M., Zhang, L., and Xu, Q. (2023). Are Transformers Effective for Time Series Forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 11121–11128.
- Zhang, C.-X., Li, J., Huang, X.-F., Zhang, J.-S., and Huang, H.-C. (2022a). Forecasting stock volatility and value-at-risk based on temporal convolutional networks. *Expert Systems with Applications*, 207(May 2020):117951.
- Zhang, J., Wang, P., Yan, R., and Gao, R. X. (2018a). Deep Learning for Improved System Remaining Life Prediction. *Procedia CIRP*, 72(January):1033–1038.
- Zhang, L., Wang, S., and Liu, B. (2018b). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4).
- Zhang, Q., Qin, C., Zhang, Y., Bao, F., Zhang, C., and Liu, P. (2022b). Transformer-based attention network for stock movement prediction. *Expert Systems with Applications*, 202(November 2020):117239.
- Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., Dewan, C., Diab, M., Li, X., Lin, X. V., Mihaylov, T., Ott, M., Shleifer, S., Shuster, K., Simig, D., Koura, P. S., Sridhar, A., Wang, T., and Zettlemoyer, L. (2022c). OPT: Open Pre-trained Transformer Language Models.
- Zhang, W., Zhao, L., Xia, H., Sun, S., Sun, J., Qin, M., Li, X., Zhao, Y., Zhao, Y., Cai, X., Zheng, L., Wang, X., and An, B. (2024). A Multimodal Foundation Agent for Financial Trading: Tool-Augmented, Diversified, and Generalist.

- Zhang, X., Liang, X., Zhiyuli, A., Zhang, S., Xu, R., and Wu, B. (2019). AT-LSTM: An Attention-based LSTM Model for Financial Time Series Prediction. *IOP Conference Series: Materials Science and Engineering*, 569(5):052037.
- Zhao, H., Liu, Z., Wu, Z., Li, Y., Yang, T., Shu, P., Xu, S., Dai, H., Zhao, L., Mai, G., Liu, N., and Liu, T. (2024). Revolutionizing Finance with LLMs: An Overview of Applications and Insights.
- Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., and Ba, J. (2023). Large Language Models Are Human-Level Prompt Engineers. In *ICLR 2023*.

Literature review

A Studies on AI in finance

Table A.1: Literature review on stock-market forecasting using AI									
Ref.	Year	Dataset	Period	Features	Lag	Horiz.	Methods	Error	
Gangopadhyay and Majumder (2023)	2023	NIFTY 50	2010-2021	OCHL, News, Tech, indic.	10d	1d	SVM, RF, MLP, LSTM, CNN, Transformer	Accuracy	
Bhandari et al. (2022)	2022	SP500	2006-2020	Macro data	-	1d	LSTM	RSE, MAPE, R^2	
Zhang et al. (2022a)	2022	88 most tweeted stocks	2014-2016	Price, Tweets	5d	1d	RNN + LSTM	Accuracy	
Wang et al. (2022)	2022	CSI, Nikkei, H. Seng, SP500	2010-2020	Daily closing prices	9d	1d	RNN, CNN, LSTM, Transformer	MAE, MSE, MAPE	
Gao et al. (2022)	2022	SP500, DJIA, NASDAQ	2018-2019	OCHLIV	1-5d	1d	SVM, RF, CNN, RNN, LSTM, GRU, 2D-CNN	Accu., prec., rec., F-meas., ROC	
Gao et al. (2021)	2021	Shanghai	2007-2021	OCHLIV, Tech, indic.	10-50d	trend	LSTM, GRU	MSE, RMSE, MAE	
Qin et al. (2020)	2020	SP500, DJIA, HSI	2000-2019	OCHLIV	-	1d	LSTM + Attention	MSE, RMSE, MAE	
Hiransha et al. (2018)	2018	NSE (5 stock)	1997-2016	OCHLIV, Trades	200d	1-10d	LSTM, RNN, CNN, MLP	MAPE	
Lee and Yoo (2020)	2018	SP500 (10 stocks)	2012-2016	OCHLIV	36m	1m	LSTM, RNN, GRU	Accuracy	
Samarawickrama and Fernando (2017)	2017	CSE (297 stocks)	2012-2013	OCHLIV	2d	1d	LSTM, RNN, GRU	MAD, MAPE	
Sethu et al. (2017)	2017	NSE (3 stocks)	2007-2017	Price	30d	1-5d	LSTM, CNN	Accuracy	
Li et al. (2017)	2017	Daily stock price	2007-2017	OCHLIV	-	-	LSTM	RMSE, MAE	
Ewa Dewi (2016)	2016	Rommania stock market	2001-2016	OCHLIV	-	1d	LSTM	Weekly return	
Heaton et al. (2017)	2017	SP500 (20 stocks)	2010-2015	Price	-	-	LSTM, AE	Accuracy	
Chen et al. (2015)	2015	China stock market	1990-2015	OCHLIV	30d	3d	LSTM	Accuracy	

Agent results

A Results 16-27 Dec'24

A.1 Prompting scenario 1

Table A.1: MSE Technical +
Fundamental + News LSTM vs
ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	3	206.92	236.11	509.14
MSFT	12.05	57.87	115.44	336.62
NVDA	115.69	163.72	141.29	176.57
TSLA	273.29	1336.95	2122.76	3034.23
META	878.12	138.32	511.26	3675.36
GOOG	23.06	24.68	39.75	29.39
AMZN	185.47	187.44	436.47	235.92

Table A.2: MSE Technical +
Fundamental + News AFAs vs
ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	2.06	3.02	14.45	0.68
MSFT	5.52	5.75	7.98	5.94
NVDA	5.25	1.81	7.71	29.43
TSLA	13.25	19.32	135.46	8.19
META	6.37	24.29	14.65	25.43
GOOG	1.95	0.83	1.77	8.04
AMZN	2.19	4.42	4.41	2.39

Table A.3: MAPE Technical +
Fundamental + News LSTM vs
ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	0.62	5.17	5.67	8.39
MSFT	0.67	1.52	2.2	3.74
NVDA	6.78	8.54	8.12	9.21
TSLA	3.22	7.37	9.31	11.26
META	3.91	1.7	3.32	8.91
GOOG	2.18	2.36	3.04	2.63
AMZN	4.94	5.32	8.31	6.15

Table A.4: MAPE Technical +
Fundamental + News AFAS vs
ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	0.52	0.55	1.41	0.28
MSFT	0.41	0.43	0.49	0.47
NVDA	1.43	0.74	1.85	3.63
TSLA	0.74	0.8	2.32	0.46
META	0.36	0.58	0.48	0.68
GOOG	0.67	0.37	0.61	1.21
AMZN	0.5	0.68	1.56	0.5

A.2 Prompting scenario 2

Table A.5: MSE Technical +
Fundamental LSTM vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	28.86	39.08	26.22	290.92
MSFT	81.64	118.42	232.7	159.09
NVDA	1.78	4.73	15.08	111
TSLA	4879.33	5397	9662.45	12768.59
META	746.78	2106.47	2746.4	3987.53
GOOG	152.45	114.09	168.91	188.32
AMZN	377.65	400.77	505.86	841.12

Table A.6: MSE Technical +
Fundamental AFAs vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	0.62	0.75	1.15	2.22
MSFT	4.79	2.06	48.03	1.31
NVDA	10.17	12.73	0.47	1.52
TSLA	23.15	347.06	43.5	22.31
META	7.51	26.6	8.94	22.97
GOOG	21.42	34.7	16.5	8.68
AMZN	1.58	2.38	1.98	30.36

Table A.7: MAPE Technical +
Fundamental LSTM vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	1.8	2.27	1.91	6.43
MSFT	1.69	2.19	3.16	2.67
NVDA	0.97	1.45	2.59	7.13
TSLA	13.11	15	20.68	24.2
META	3.75	6.66	7.79	9.49
GOOG	5.34	5.02	6.29	6.77
AMZN	7.01	7.86	9.1	11.96

Table A.8: MAPE Technical +
Fundamental AFAs vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	0.28	0.32	0.39	0.53
MSFT	0.43	0.3	1.35	0.21
NVDA	2.27	2.51	0.43	0.74
TSLA	0.88	4.02	1.19	0.94
META	0.42	0.7	0.4	0.62
GOOG	2.02	2.75	1.94	1.21
AMZN	0.44	0.64	0.47	2.11

A.3 Prompting scenario 3

Table A.9: MSE News LSTM vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	56.42	95.2	89.2	119
MSFT	166.62	225.59	207.09	207.09
NVDA	7.41	0.66	1.07	35.98
TSLA	5155	5732.63	7116.28	5956.6
META	945.33	1937.97	2047.76	785.84
GOOG	148.56	132.7	140.23	217.43
AMZN	371.47	330.22	674.97	726.46

Table A.10: MSE News AFAs vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	8.96	8.98	16.58	42.75
MSFT	48.84	33.73	42.29	42.29
NVDA	3.98	0.77	5.12	2.01
TSLA	47.11	264.87	28.29	23.09
META	8.36	5.48	61.91	22.97
GOOG	19.98	17.25	16.95	29.78
AMZN	7.17	20.82	25.88	62.63

Table A.11: MAPE News LSTM vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	2.5	3.53	3.52	4.15
MSFT	2.39	3.02	2.98	2.98
NVDA	1.85	0.54	0.7	3.96
TSLA	13.47	15.46	17.74	16.55
META	4.2	6.39	6.74	4.24
GOOG	5.27	5.41	5.73	7.28
AMZN	6.96	7.14	10.51	11.24

Table A.12: MAPE News AFAs vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	1.09	1.12	1.39	2.74
MSFT	1.48	1.22	1.23	1.23
NVDA	1.17	0.57	1.38	0.84
TSLA	1.38	3.45	0.96	0.96
META	0.39	0.29	1.03	0.62
GOOG	2.33	2.04	2.01	2.6
AMZN	1.07	1.85	1.81	2.81

B Results 23 Dec'24-3 Jan'25

B.1 Prompting scenario 1

Table B.1: MSE Technical + Fundamental + News LSTM vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	13.73	71.93	65.66	71.79
MSFT	31.11	13.24	28.99	54.12
NVDA	15.11	3.26	10.82	5.86
TSLA	1917.86	4546.82	2989.33	7072.04
META	-	-	-	-
GOOG	70.9	29.18	38.35	16.89
AMZN	48.86	87.65	133.28	426.05

Table B.2: MSE Technical + Fundamental + News AFAS vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	7.46	10.69	13.74	34.04
MSFT	17.47	27.73	26.09	131.68
NVDA	0.12	7.1	5.2	7.34
TSLA	23.37	127.94	30.82	650.09
META	-	-	-	-
GOOG	3.78	2.16	3.93	8.84
AMZN	7.97	1.89	12.05	26.2

Table B.3: MAPE Technical + Fundamental + News LSTM vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	1.32	3.06	2.99	3.18
MSFT	1.11	0.78	1.17	1.62
NVDA	2.41	1.21	2.26	1.68
TSLA	8.72	14.44	12.04	18.77
META	-	-	-	-
GOOG	3.63	2.54	2.99	2.03
AMZN	2.7	3.81	4.8	8.67

Table B.4: MAPE Technical + Fundamental + News AFAs vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	0.88	1.15	1.3	2.06
MSFT	0.96	1.11	1.06	2.47
NVDA	0.22	1.74	1.37	1.6
TSLA	0.93	2.13	1.07	5.36
META	-	-	-	-
GOOG	0.87	0.63	0.88	1.43
AMZN	1.18	0.54	1.42	2.08

B.2 Prompting scenario 2

Table B.5: MSE Technical +
Fundamental LSTM vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	62.05	142.18	28.24	313.55
MSFT	22.19	27.2	2.05	8.03
NVDA	5.61	0.26	19.37	16.4
TSLA	2057.52	2176.66	3289.15	5005.67
META	253.3	501.09	264.7	67.57
GOOG	34.72	71.64	70.82	37.76
AMZN	30.04	142.36	253.83	96.76

Table B.6: MSE Technical +
Fundamental AFAs vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	0.16	13.02	7.92	42.37
MSFT	15.6	47.27	14.56	11.43
NVDA	0.12	2.06	3.78	0.97
TSLA	17.07	10.92	14.51	913.46
META	2.8	34.39	481.93	1341.39
GOOG	2.56	1.34	10.77	10.76
AMZN	4.54	0.71	1.65	1.65

Table B.7: MAPE Technical +
Fundamental LSTM vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	2.67	4.26	1.98	6.57
MSFT	0.95	1.11	0.28	0.6
NVDA	1.48	0.25	3.02	2.83
TSLA	9.03	10.01	12.62	15.78
META	2.25	3.42	2.55	1.2
GOOG	2.56	3.96	4.05	3.02
AMZN	2.14	4.84	6.6	4.09

Table B.8: MAPE Technical +
Fundamental AFAs vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	0.11	1.22	0.93	2.26
MSFT	0.84	1.52	0.83	0.76
NVDA	0.22	0.82	1.18	0.58
TSLA	0.94	0.66	0.65	6.02
META	0.25	0.81	3.12	5.46
GOOG	0.79	0.56	1.49	1.47
AMZN	0.93	0.32	0.5	0.5

B.3 Prompting scenario 3

Table B.9: MSE News LSTM vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	38.35	49.69	25	10.08
MSFT	4.27	0.34	13.68	26.71
NVDA	1.15	0.23	19.94	20.77
TSLA	3824.78	2404.87	2948.36	3175.39
META	351.38	103.83	96.28	39.51
GOOG	22.59	31.86	81.48	16.89
AMZN	25.46	174.19	136.59	169.71

Table B.10: MSE News AFAs vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	8.98	3.96	6.91	96.25
MSFT	12.36	8.36	22.94	37.11
NVDA	1.81	2.03	48.09	5.02
TSLA	3.61	3.41	30.75	308.11
META	23.32	192.95	341.13	539.4
GOOG	4.38	7.74	21.52	8.77
AMZN	26.18	55.71	35.27	60.32

Table B.11: MAPE News LSTM vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	2.13	2.55	1.87	1.21
MSFT	0.44	0.11	0.8	1.14
NVDA	0.67	0.26	3.06	3.18
TSLA	12.24	10.51	11.89	12.48
META	2.64	1.56	1.52	0.85
GOOG	2.08	2.66	4.34	2.03
AMZN	1.82	5.2	4.8	5.4

Table B.12: MAPE News AFAs vs ARIMA

Ticker	PX_{d+3}	PX_{d+5}	PX_{d+7}	PX_{d+10}
AAPL	1.09	0.71	0.93	3.51
MSFT	0.77	0.56	0.94	1.28
NVDA	0.9	0.79	4.98	1.32
TSLA	0.35	0.35	1.15	3.45
META	0.65	2.17	2.48	3.42
GOOG	0.96	1.28	2.24	1.29
AMZN	2.22	3.31	2.56	3.37