



Exploring the Interplay of Quarterly Financial Metrics, Multi-Agent Sentiment Analysis, and Short-Term Stock Market Trends

<i>Student</i>	<i>Shvan Jaro</i>
<i>Administration Number</i>	<i>779681</i>
<i>Student Number</i>	<i>2108749</i>
<i>Telephone Number</i>	<i>+31657946208</i>
<i>Email Address</i>	<i>s.jaro@tilburguniversity.edu</i>
<i>Study Programme</i>	<i>MSc Information Management</i>
<i>Faculty</i>	<i>Tilburg School of Economics and Management</i>
<i>Supervisor</i>	<i>Dr. Ir. Francesco Lelli</i>
<i>Second Reader</i>	<i>Dr. Shiva Shekhar</i>

Tilburg University

Warandelaan 2

5037 AB, Tilburg

Netherlands

Year 2023 – 2024

06-06-2024

(0.i) Abstract

The relationships between quarterly financial earnings metrics, multi-agent sentiment analysis, and short-term stock market trends are crucial yet underexplored. This thesis addresses this gap by combining quantitative analysis of financial metrics with qualitative sentiment analysis using advanced Artificial Intelligence (AI) models, specifically Llama 8B Instruct and GPT-3.5. By examining data from nine S&P 500 companies over the 2022-2023 period, the study uncovers significant correlations between financial performance, market sentiment, and short-term stock trends, particularly around earnings release dates. The methodology involves detailed quantitative analysis of the variables and qualitative sentiment analysis of news headlines and tweets, leveraging the capabilities of state-of-the-art AI models combined with the innovative concept of multi-agent conversations. This dual approach allows for a nuanced understanding the interplay of financial metrics, market sentiment, and short-term price trends. Key findings highlight that analysing earnings metrics and their relationship to sentiment offers deeper potential insights into market dynamics.

Key Words: Quarterly financial metrics, Company earnings, Large Language Models, Artificial intelligence in finance, Artificial Intelligence agents, Multi-Agent systems, LLM collaborations

(0.ii) Preface

This thesis marks the end of my academic journey, representing the final artifact of my educational deliverables. Conducted the past half year, it was filled with challenges and triumphs. Being at the forefront of innovation brought its own set of challenges. However, these experiences taught me that with hard work and determination, solutions are always within reach. I would like to express my gratitude to the two most important women in my life, my mother and girlfriend. Who've always believed in me and supported me tremendously in the past half year. I would also like to express my sincere appreciation to Dr. Ir. Francesco Lelli. Your guidance and relentless push for excellence have inspired me to strive for the best in myself. Your mentorship has been invaluable, and I am profoundly thankful for your support throughout this journey. I will miss the fun supervision sessions.

Bergen op Zoom, NL, 06/06/2024.

(0.iii) Addressing the AI tools used

In accordance with the school's policy on transparency regarding the use of AI tools, I am dedicating this section to acknowledge and detail how I have utilized such tools in the development of this thesis. Throughout the research and writing process, I have used multiple AI tools as both writing assistants and thinking aids. Specifically, I employed the following:

- *GPT-4/4o (OpenAI, n.d.-a)*: I used GPT-4/4o mostly for writing and sparring sessions. Whenever I encountered concepts that I did not fully understand, I engaged in discussions with GPT-4/4o to gain clarity and deepen my understanding. This tool proved invaluable in helping me articulate my ideas clearly and effectively.
- *Google Gemini (Google, n.d.)*: I primarily used Google Gemini as a replacement for traditional Google search. This model helped me find relevant academic papers, understand complex and lengthy texts, and gather comprehensive information on various topics related to my research.

The use of these AI tools has been important in enhancing the quality and depth of my work. They provided me with additional perspectives, facilitated my understanding of intricate concepts, and assisted in the presentation of my findings.

(0.iv) Index

1	Introduction	8
1.1	Problem Indication	8
1.2	Problem Statement	10
1.3	Thesis structure	11
2	Theoretical Overview.....	12
2.1	Sentiment, Data, and Market Dynamics.....	12
2.2	Company Reports.....	12
2.3	Natural Language Processing	13
2.4	Sentiment Analysis.....	14
2.4.1	Sentiment analysis in Finance.....	15
2.5	The Transformer Architecture	15
2.6	Generative AI.....	16
2.7	Transformer model and Sentiment Analysis	17
2.8	Multi-Agent Conversation Framework	18
2.8.1	AutoGen.....	20
2.9	Prompt Engineering.....	22
2.10	Conceptual Model	24
2.11	Summary.....	25
3	Quarterly Financial Metrics, News, Tweets, and Price Data	26
3.1	Data Specifications.....	26
3.1.1	Quantitative data (Q-data)	27
3.1.2	News headlines and Tweets (QI-Data).....	28
3.1.3	Price data	29
3.1.4	Data Collection, Cleaning, And Homogenization.....	30
3.1.5	Code and Tools Used	31
3.1.6	Merged Dataset	31
4	Experimental Research Design.....	32
4.1	Experimental Design	33
4.2	Defining the Agents.....	34
4.3	Multi-Agent Sentiment Analysis: Conversation Set-up.....	36
4.3.1	Group Chat Workflow.....	37
4.3.2	Tailoring the Prompts (Human Prompt).....	39
4.3.3	Tailoring the Prompts (Agent Messages)	41
4.4	Sentiment Analysis.....	42

4.5	Price Trend Definition	42
4.6	Statistical Analysis	45
4.7	Declaration of Competing Interest	46
5	Results	47
5.1	Correlation Analysis	47
5.2	Financial Metrics and Stock Trends	48
5.3	Multi-Agent Sentiment Analysis and Stock Price Trends	49
5.4	Financial metrics and Multi-Agent Sentiment	50
6	Discussion	54
6.1	Addressing the Research Questions	54
6.2	Limitations	57
6.3	Recommendations for Future Research	58
7	Conclusion	60
	References	61
	Appendix A: Data	67
A.1	Data specifications table	67
A.2	Data Value	68
A.3	Metrics used in Q-file	69
A.4	Quantitative Data File Snapshot	71
A.5	Qualitative Data File Snapshot	72
A.6	Final Data File Snapshot	73
	Appendix B: Python code	74
B.1	Quantitative Data Homogenization	74
B.1.1	Data Conversion Steps	74
B.2	Qualitative Data Cleaning	75
B.3	Moving Average Calculations	77
B.4	Determining the Price Trends	79
B.5	Multi-Agent Conversation Framework	83
B.5	Correlation Matrix	89
	Appendix C: Prompt Design	91
C.1	Prompt Design Words Analyst	91
C.2	Prompt Design Sarcasm Analyst	92
C.3	Prompt Design Emotion Analyst	93
C.4	Prompt Design Sentiment Analyst	95
C.5	Prompt Design Sarcasm Analyst	98

Appendix D: Example of Multi-Agent Chats.....	101
D.1 Result of tests with GPT 3.5	101
D.2 Result of tests with Llama 8B Instruct	105

1 Introduction

In this chapter, the context, relevance, and objectives of the study are described. Starting with the problem indication, it sets the frame for the research, justifying the domain for the study. The problem statement then shapes this broad context into the intent of the research, identifying the knowledge gaps. The research question section breaks down the problem into elements that can be systematically addressed. Finally, the thesis structure gives an organizational layout of the thesis, explaining how each part will build upon the other to unfold the narrative of the research.

1.1 Problem Indication

The conventional view is that movements in asset prices are primarily driven by changes in their intrinsic values. However, financial markets are complex systems influenced by a multitude of factors, including company earnings and market sentiment (Cutler et al., 1988). Earnings announcements are the main way public companies update investors on their periodic financial performance (Dechow et al., 2013). Consequently, it's unsurprising that a large amount of research focuses on the relationship between stock prices and earnings. The relationship between company earnings and stock prices encompasses a range of topics, from the impact of earnings report timing on stock price reactions (Chambers & Penman, 1984) to the integration of both qualitative and quantitative aspects of quarterly earnings reports in predicting stock price movements (Lin et al., 2011). Beyond these quantitative factors, sentiment also plays an important role in financial markets. Sentiment can be defined in various ways, but in this research, the focus of sentiment analysis is on studying people's opinions, evaluations, attitudes, and emotions toward organizations (Liu, 2012). Sentiment analysis primarily draws from the following three sources: public and corporate disclosures/filings, media articles, and internet messages (Kearney & Liu, 2013). The growing influence of sentiment is underscored by the widespread use of the internet, with more than 66 % of the global population now online. Recent data shows a total of 5.35 billion internet users worldwide, a number that has increased by 1.8 percent over the past year, adding 97 million new users since the beginning of 2023 (Thompson, 2024). This vast digital presence provides a rich source of data for sentiment analysis, reflecting the collective mood and opinions of investors and the general public. Capturing sentiment is an actively researched topic in Natural Language Processing (NLP). Since the introduction of the transformer architecture by Vaswani et al., (2017), the field of NLP has experienced a paradigm shift, leading to significant

improvements in AI's ability to understand and generate human-like text. Deep learning has become the leading approach for sentiment analysis, surpassing traditional machine learning methods (Abdullah & Ahmet, 2022). The introduction of the transformer architecture offers significant opportunities for use in capital markets and has been leveraged in various applications in capital markets. For example, previous research utilized sentiment analysis conducted by AI for predicting prices (Jaiswal et al., 2023; Sidogi et al., 2021). Integrating advanced AI techniques with traditional financial models can lead to more sophisticated and effective solutions in the financial industry. The application of AI in financial markets has a long history, ranging from analyzing market data and financial news to identify trading opportunities, to aiding in the selection of the best portfolio of assets to maximize returns and minimize risks (Koshiyama et al., 2020). With the transformer architecture, enhanced Large Language Models (LLMs) were developed. These LLMs conduct pre-training on large amounts of text data and employ various training techniques (Zhang et al., 2023). Building on this foundation, the rapid innovation in AI within the finance sector has introduced new concepts that are transforming the field. One such advancement is the use of LLM Multi-Agent Systems. These systems leverage collective intelligence for complex problem-solving and world simulation, enabling multiple agents to collaborate effectively in tackling intricate financial problems (Guo et al., 2024). LLM Multi-Agent systems have proven to outperform single LLMs in sentiment analysis tasks (Sun et al., 2023).

The focus of this study is to uncover the relationships between quarterly financial earnings metrics, multi-agent sentiment analysis, and short-term stock market trends. I will examine how quarterly financial metrics, as published in company reports, correlate with short-term stock trends. Additionally, the study emphasizes analyzing sentiment using multi-agent conversations to explore the relationships between sentiment analysis and short-term market trends, employing LLMs as key components. The multi-agent sentiment analysis is conducted on news headlines and tweets. The news headlines and tweets are collected in the timespan of quarterly financial metrics releases, to investigate the impact of these financial metrics on market sentiment. This research encompasses both quantitative and qualitative aspects, including quantitative quarterly metrics and qualitative data from news headlines and tweets, upon which the sentiment analysis is performed.

1.2 Problem Statement

While sentiment analysis and company reports have been extensively researched, several knowledge gaps persist. Research on company reports includes topics such as timeliness (Chambers & Penman, 1984), investor perceptions of annual earnings announcements and their informational value (Beaver, 1968), and the impact of persistent earnings innovations on stock returns (Kormendi & Lipe, 1987). However, the relationships between specific financial earnings metrics in these reports and short-term stock trends remain underexplored. Although Lin et al., (2011) examined both quantitative and qualitative aspects of financial reports, their focus was primarily on the qualitative aspects, analyzing sentiment with four financial metrics. The use of transformer architecture for sentiment analysis has been widely adopted, with BERT (Bidirectional Encoder Representations from Transformers) being the most commonly utilized deep learning model for these tasks (Devlin et al., 2019; Jaiswal et al., 2023; Sidogi et al., 2021). LLMs such as GPT have demonstrated better performance than BERT in sentiment analysis (Wang et al., 2024). However, utilizing LLM negotiations, where there is collaboration, has been found to outperform traditional methods (Sun et al., 2023). While the performance of LLM negotiations has been researched, the relationships of multi-agent sentiment analysis and short-term stock is a remaining gap in current research. The outcome of this study will be an exploratory statistical analysis of the variables. It will address relationships between specific quarterly financial metrics and short-term stock trends. It will also examine the relationship between multi-agent sentiment analysis and short-term stock trends. Additionally, the study will explore the connection between quarterly financial metrics and market sentiment. Examining these relationships will provide insights into how financial performance and market sentiment relate to short-term stock trends, and the relationship of the financial metrics with market sentiment. This analysis may lead to a deeper understanding of market dynamics in conjunction with innovative tools. Building on this, I ask the following research questions:

- *RQ1*: “How do specific quarterly financial metrics relate to stock market trends?”
- *RQ2*: “How does multi-agent sentiment analysis of news headlines and tweets relate to stock market trends?”
- *RQ3*: “What are the relationships between financial metrics and multi-agent sentiment around the time of quarterly financial report releases?”

These research questions aim to explore the relationship between qualitative sentiment data from news headlines and tweets, quantitative quarterly financial data, and stock market trends. To answer these questions, a mixed-method approach of quantitative and qualitative analysis is employed. Quantitative methods, correlation analysis, are used to assess correlations. Qualitative methods, such as contextual and content analysis, provide deeper insights into the nuances of sentiment and its relationship with market movements. Quarterly data is collected for nine companies listed in the Standard & Poor's 500 (S&P 500) Index, namely Apple, Amazon, Alphabet (Google), Berkshire Hathaway, Eli Lilly, Meta, Microsoft, Nvidia, and Tesla.

1.3 Thesis structure

This thesis unfolds across seven chapters, exploring the interplay between multi-agent sentiment analysis, quarterly financial earnings data, and short-term stock market trends. The structure of the thesis is as follows:

- *Chapter 1: Introduction* - Outlines the motivation, problem statement, and research questions guiding the goal of the study.
- *Chapter 2: Theoretical Overview* - Lays the foundation by reviewing literature on financial markets, company reports, sentiment analysis, NLP, transformer models, LLMs, and multi-agent conversations.
- *Chapter 3: Data Chapter* - Focuses on the data collection process, detailing how social media, news, and financial report data are accessed and obtained through Refinitiv Workspace.
- *Chapter 4: Experimental Setup* - Describes the methodology and experimental setup, discussing the steps taken to conduct the analysis.
- *Chapter 5: Results* - Presents the findings of the analysis.
- *Chapter 6: Discussion* - Evaluates the implications of these findings within the broader context, identifies limitations, and suggests directions for future research.
- *Chapter 7: Conclusion* - Synthesizes the key insights and contributions of the research.

By methodically exploring these chapters, this thesis aims to provide a understanding of the relationships between multi-agent sentiment analysis, quarterly financial earnings data, and short-term stock market trends.

2 Theoretical Overview

This chapter establishes an overview to understand the integration of financial markets, investment theories and company reports with the concept of AI. Beginning the chapter, financial markets and company reports are introduced. Following AI, sentiment analysis, and the multi-agent conversation framework are discussed, utilizing the AutoGen library. After this, the concept prompt engineering is explored, a method to program AI through prompts. The chapter concludes with the conceptual model, formed based on the knowledge gaps found in the theoretical overview.

2.1 Sentiment, Data, and Market Dynamics

Combining the insights from the Efficient Market Hypothesis (EMH), behavioral finance, fundamental analysis, and the role of social media in financial markets is foundational for this research. The EMH posits that all publicly available information is already reflected in stock prices, limiting the potential for higher returns through information-based strategies (Fama, 1970). However, a vast amount of financial data is now available to investors and stakeholders. This data is available online in various formats, both qualitative (textual) and quantitative (numerical) (Magnusson et al., 2005). The introduction of social media introduces a dynamic layer of investor sentiment and perception. This real-time availability of public sentiment can precede formal financial disclosures and analyses. Bukovina (2016) highlights that this challenges the EMH's assertion of market efficiency in the digital age. Behavioral finance enriches this discussion by underscoring the significant impact of psychological factors and biases on investor decisions and market outcomes. Irrational behaviors and cognitive biases identified by behavioral finance explain market anomalies and inefficiencies. These anomalies are often amplified by the rapid spread of information and sentiment on internet platforms (Barberis & Thaler, 2003; Shleifer & Summers, 1990).

2.2 Company Reports

A company's financial report is an essential measure of financial performance, offering accounting items and financial ratios (Magnusson et al., 2005). Earnings announcements are the main events through which companies update investors on their periodic performance. Since 1970, publicly traded companies in the United States have been mandated to submit quarterly income statements to the Securities and Exchange Commission via Form 10-Q

(Dechow et al., 2013). Researchers have thoroughly examined the concepts of quarterly reports in financial markets. Chambers and Penman, (1984) investigate the impact of the timeliness of earnings reports on stock price reactions. They find that earlier-than-expected reports tend to be associated with positive abnormal returns, while delayed reports often carry negative news. Magnusson et al. (2005) examine the language of quarterly reports as an indicator of changes in a company's financial status. They emphasize the importance of accounting items and financial ratios in assessing financial performance. Lin et al., (2011) show that incorporating quarterly financial reports for predicting short-term price movements enhances prediction accuracy and profitability. They achieve this by effectively combining qualitative and quantitative features through advanced clustering techniques. With timeliness, qualitative, and quantitative aspects being addressed in current research, there remains a knowledge gap in the comparing the relationship of specific quantitative metrics of quarterly reports with stock market trends. According to Beaver, (1968), annual earnings reports have substantial informational value, influencing both the volume and price movements of common stocks in the weeks surrounding the announcement. This highlights the importance of earnings announcements in shaping investor behavior and market dynamics.

Kormendi & Lipe (1987) explain that before earnings announcements, investors form expectations based on available information and past performance, engaging in speculative activities. On the day of the announcements, the market reacts immediately, with stock prices adjusting significantly based on the new earnings information. After the announcements, the impact of these innovations persists as the market continues to process the information. This ongoing adjustment reflects the concept of earnings persistence, where the influence of earnings innovations on stock returns extends beyond the initial announcement period.

2.3 Natural Language Processing

AI, a branch of computer science, is mainly concerned with automation of intelligent behavior. In the dynamic intersection of grammar and AI, NLP emerges as a field dedicated to bridging the gap between human language understanding and machine interpretation (Chowdhary, 2020). Chowdhary, (2020) outlines the foundational concepts of AI as follows:

- *Data Structures*: These are methods for organizing and storing data efficiently, enabling quick access and modification.

- *Knowledge Representation Techniques*: These methods allow AI to produce information about concepts in a form that a computer system can utilize to solve difficult objectives.
- *Algorithms*: These are step-by-step procedures or formulas for solving problems and making decisions by applying the represented knowledge.
- *Programming Techniques*: These are the methods and practices used to implement the data structures, knowledge representations, and algorithms in software to build functional AI systems.

As articulated by Chowdhary (2020), languages, whether bounded by a finite set of symbols or envisioned as infinite arrays of sentences, present a complex domain for computational analysis. NLP seeks to navigate this complexity by employing computational techniques for the automatic analysis and representation of human language, motivated by foundational linguistic theory.

2.4 Sentiment Analysis

Sentiment analysis is a branch of NLP that explores how people perceive and react to different entities such as events, topics, services, products, and organizations. It aims to decode feelings, attitudes, emotions, and opinions, along with their specific attributes (Liu, 2012). Also known as opinion mining, sentiment analysis aims to analyze collected data ranging from text snippets and sentences to entire documents to understand user sentiments. This process assigns a score that reflects the sentiment detected in the data, offering insights into the users' perspectives and emotions (Singh et al., 2024). The evolution entails from neural network architectures in sentiment analysis, to recurrent and convolutional neural networks, to the transformative impact of transformer models like BERT, GPT, and others (Abdullah & Ahmet, 2022). Sentiment analysis can be categorized into the following three main levels (Liu, 2012).

- *Document level*: Classifies the overall sentiment of an entire document as positive or negative. For example, determining if a product review is overall positive or negative.
- *Sentence level*: Determines the sentiment of each sentence, classifying it as positive, negative, or neutral. Identifying the sentiment expressed in each sentence of a customer review.

- *Entity and Aspect Level*: Provides a finer-grained analysis by identifying sentiments on specific entities and their aspects. Focuses on the opinion targets, offering detailed insights into what aspects people like or dislike. For example, Analyzing the sentence "The iPhone's call quality is good, but its battery life is short" to determine positive sentiment on call quality and negative sentiment on battery life.

2.4.1 Sentiment analysis in Finance

AI and Machine Learning are reshaping the financial industry (Koshiyama et al., 2020). Applying sentiment analysis to capital markets, a growing interest has been emphasized, due to its ability to support business decision-making and financial forecasting (Du et al., 2024). With this, the research on Financial Sentiment Analysis (FSA) can be divided into various specific types of sentiment analysis studies. Du et al., (2024) proposes the two forms of FSA:

- *Technique-Driven FSA Studies*: Technique-driven FSA studies primarily focus on the sentiment analysis of financial texts. The main objective is to develop and refine techniques and models tailored for FSA. These studies emphasize detecting sentiment granularity, such as targeted aspect-based sentiment analysis or sentence-level analysis, to enhance accuracy and specificity.
- *Application-Driven FSA Studies*: Application-driven FSA (Financial Sentiment Analysis) studies focus on utilizing financial sentiment as a proxy for investor sentiment to predict movements in financial markets. The objective of these studies is to explore various financial application scenarios, including predicting stock market movements, financial risk, and movements in FOREX and cryptocurrency markets. These studies leverage datasets that are annotated by financial metrics derived from market data, typically formatted as time series. These datasets include extensive financial texts and cover representative periods essential for modeling the relationships between investor sentiment and financial metrics.

2.5 The Transformer Architecture

The transformer architecture, introduced by Vaswani et al., (2017), has changed the state-of-the-art for LLMs. Unlike older methods that processed words one by one (RNNs and CNNs) the transformer model looks at the whole sentence at once. The state-of-the-art shifted to transformer models because it enhances the speed of training and allows the model to get a

better sense of context and relationships between words (Vaswani et al., 2017). This model is made up of two main parts: one that encodes (breaks down and understands) the input text and another that decodes (builds up) the output text. Both parts use “self-attention”, which helps the model pay attention to important words in the sentence, regardless of the position of the words in the sentence (Vaswani et al., 2017). The advent of the transformer architecture has catalyzed various research across various domains. Notably in finance and stock market trend prediction with sentiment analysis (Jaiswal et al., 2023; Sidogi et al., 2021; Song et al., 2023).

2.6 Generative AI

LLMs are advanced computational systems designed to comprehend and produce human language. They possess the transformative capability to predict the probability of word sequences and generate new text from a given input (Chang et al., 2024). LLMs, such as the Generative Pre-trained Transformer (GPT), have recently achieved notable progress due to the advent of the transformer architecture. These models can produce text that resembles that of a human, respond to queries, and carry out other language-related activities with great accuracy since they have been trained on vast volumes of text data (Kasneci et al., 2023). LLMs can use this training data to generate novel outputs, ranging from text to images, that resonate with the training data's structure and essence. It provides a mechanism to produce synthetic financial narratives or predictive text that can enrich the dataset for training more robust models (Brynjolfsson et al., 2023).

Recent advances in LLMs have introduces new possibilities for their applications in finance. LLMs can be utilized for various financial tasks through methods such as zero-shot or few-shot learning, fine-tuning on domain-specific data, and even training custom LLMs from scratch (Li et al., 2023). These techniques have shown significant performance improvements in financial NLP tasks like sentiment analysis, news summarization, and financial risk modelling. For instance, Li et al., (2023) states the following applications:

- *Zero-shot and Few-shot Learning*: LLMs can perform tasks with little to no task-specific training data. This is particularly useful in finance where annotated data may be scarce. For example, LLMs can classify financial sentiment or summarize news articles with minimal examples provided.

- *Fine-tuning on Domain-specific Data:* Fine-tuning pre-trained LLMs on financial datasets enhances their ability to understand and generate finance-specific content. This method improves performance on tasks such as financial report analysis and risk assessment.
- *Training Custom LLMs:* Developing models from scratch with a mix of general and finance-specific data can lead to highly specialized LLMs. These models can outperform general-purpose LLMs on specific financial tasks, offering more accurate and contextually relevant outputs.

This approach enables financial entities to start with low-cost experimentation and progress to more substantial investments in customized LLMs, ensuring that the chosen solution aligns with their specific requirements and resource constraints.

2.7 Transformer model and Sentiment Analysis

The transition to pre-trained transformer models marks a pivotal evolution in sentiment analysis. These models capitalize on their capacity to transfer refined language comprehension across diverse tasks (Abdullah & Ahmet, 2022). Du et al. (2024) explains sentiment analysis methodologies, underscoring the superiority of Pre-Trained Language Models (LLMs) for such applications. Due to their proven effectiveness, LLMs have emerged as the leading method for cutting-edge sentiment analysis. However, Zhang et al., (2023) discuss several ongoing challenges in sentiment analysis despite advancements in Pre-Trained Language Models. While LLMs have achieved success in binary sentiment classification, sentiment analysis has not yet reached maturity. Zhang et al., (2023) discusses the following challenges:

- *Understanding Complex Linguistic Nuances and Cultural Specificity:* Sentiment often involves subtleties such as sarcasm, irony, humour, and cultural idioms. Models struggle to understand these nuances without context and shared cultural knowledge.
- *Extracting Fine-Grained and Structured Sentiment Information:* Generating structured fine-grained information, such as in aspect-based sentiment analysis (ABSA), remains challenging. Such information is crucial for summarizing large-scale data and producing organized digests.

- *Real-Time Adaptation for Evolving Sentiment Analysis*: Sentiments and expressions continuously evolve, especially on social media, with new idioms and sentiment-carrying expressions emerging regularly. Current LLMs lack flexibility in fine-tuning or re-training, limiting their ability to keep pace with the fast evolution of language and sentiment. Developing methods for rapid and effective model updates is a critical research direction to ensure accurate real-time sentiment analysis.

Zhang et al., (2023) emphasize the need for sophisticated prompt engineering techniques to overcome these challenges. These techniques can better guide LLMs in recognizing and interpreting subtle cues that denote sentiment in financial discourse. Zhang et al., (2023) suggest that LLMs hold tremendous potential for revolutionizing sentiment analysis in finance. However, achieving optimal performance requires targeted adjustments and enhancements tailored to the unique challenges of financial sentiment analysis.

2.8 Multi-Agent Conversation Framework

The invent of transformer models, LLMs, and generative AI has paved the way for the emergence of Large Language Model Multi-Agent Systems. These systems enhance the capabilities of agents handling real-world tasks that require reasoning, tool usage, and rapid adjustment to new data (Wu et al., 2023). The increasing complexity of tasks drives current research into the use of many cooperative agents, leveraging the strength of collaboration within LLMs to improve performance (Guo et al., 2024).

These multi-agent systems are not just collections of independent entities; they interact through multi-agent conversations, facilitating task segmentation and integration (Wu et al., 2023). This collaborative environment significantly improves decision-making and reasoning for complex domain applications. In contrast to single-agent setups, multi-agent systems offer a robust approach for tackling dynamic and intricate tasks (Guo et al., 2024). They demonstrate the power of collective intelligence and the utilization of specialized skills, making them invaluable across a wide range of applications. The following communication paradigms between LLM Multi-Agents can be perceived (Guo et al., 2024):

- *Cooperative*: Agents collaborate towards shared goals, exchanging information to enhance collective solutions.

- *Debate*: Agents engage in argumentative interactions, presenting and defending viewpoints to reach a consensus.
- *Competitive*: Agents pursue their own goals, which may conflict with those of other agents.

Multi-agent systems utilize various communication paradigms to enhance collective intelligence, including cooperative, debate, and competitive interactions (Guo et al., 2024). Alongside these paradigms, multi-agent systems can adopt different communication structures. A visual representation is presented in Figure 1.

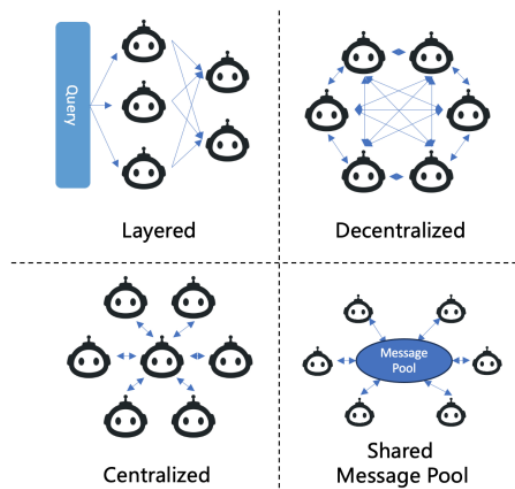


Figure 1: Communication structures of multi-agent systems (Guo et al., 2024)

The specific communication structures as specified by Guo et al., (2024) translate to the following definitions:

- *Layered Communication*: A hierarchical structure where agents have distinct roles and interact within their layer or adjacent layers.
- *Decentralized Communication*: A peer-to-peer network for direct agent communication.
- *Centralized Communication*: A central agent coordinates system communication.
- *Shared Message Pool*: Agents publish and subscribe to messages based on profiles, enhancing efficiency.

As described by Wu et al., (2023), multi-agent debate improves reasoning and factual accuracy. It is found that the debate approach outperforms single model baselines such as

zero-shot chain of thought prompting, and reflection on a variety of tasks based on reasoning, factuality, and question-answering tasks (Du et al., 2023). Figure 2 illustrates the accuracy comparison between single models and multi-agent debate approaches across various tasks. It shows that the multi-model debate consistently outperforms the single model in the categories specified (Du et al., 2023). Due to the outperforming results of multi-agent systems, the applications have crossed various domains. Ranging from software development teams to creating simulated financial trading environments, modeling agents with endowments, information, and predefined preferences to explore economic scenarios (Guo et al., 2024). However, the knowledge gap persists in using multi-agent collaboration for sentiment analysis and its relationship with short-term stock trends.

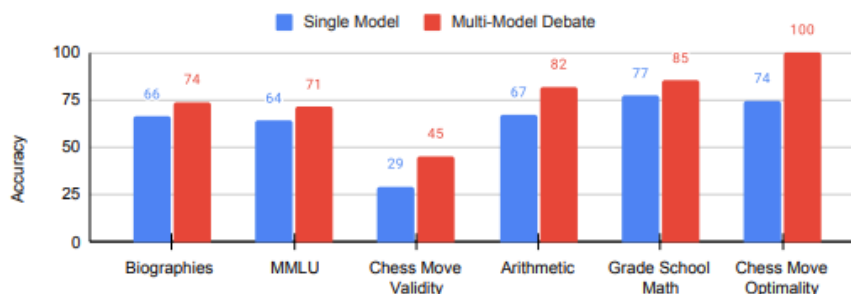


Figure 2: Accuracy Comparison Single Model (Blue) and Multi-agent Debate (Red) (Du et al., 2023)

2.8.1 AutoGen

Wu et al., (2023) introduce AutoGen, a library for facilitating multi-agent conversations. AutoGen is grounded in innovative principles of customizable and conversable agents, along with the concept of conversation programming. This framework leverages an agent architecture that integrates tools, LLMs, and human inputs. Agents within AutoGen are designed for active participation in dialogues, capable of initiating, receiving, and reacting to communications. AutoGen streamlines complex LLM application workflows through the orchestration of multi-agent dialogues. The library supports the development of LLM applications using multiple agents that converse with each other to solve tasks, offering customizable and conversable agents that integrate LLMs, tools, and human inputs.

In computational linguistics, the concept of multi-agent conversations significantly expands the utility of LLMs. AutoGen's foundation lies in agent customization, where individual conversable agents are tailored with unique capabilities or knowledge domains. These agents

can operate both independently and collectively. AutoGen supports a range of conversation patterns, offering flexibility in dialogue structure and enabling group discussions that can involve human input (Wu et al., 2023). The use cases of agents in AutoGen are illustrated in Figure 3, represented by icons that symbolize different functions. These use cases are formulated as follows (Wu et al., 2023):

- *Agent Customization*: The agents can operate both independently and collectively.
- *Flexible Conversation Patterns*: AutoGen agents are conversable, customizable, and can be based on LLMs, tools, humans, or even a combination of them.
- *Multi-Agent Conversations*: The agents have the ability to assemble into group discussions, which can involve human input.
- *Example Agent Chat*: The framework is designed to support a range of conversation patterns, offering flexibility in dialogue structure.

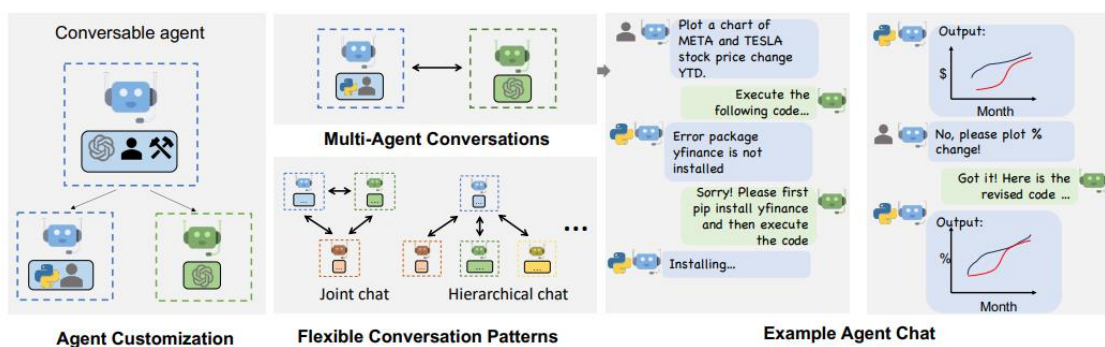


Figure 3 AutoGen Framework Use Cases (Wu et al., 2023)

The AutoGen library offers the following modules (Microsoft, 2024):

- *Agentchat*: This module is the framework designed for creating and managing agents capable of communicating with each other within AutoGen.
- *Cache*: This module outlines a protocol for basic cache operations, which is used to store data that can be retrieved quickly (e.g., previous chats with agents)
- *Code execution*: This module provides a structured approach to identifying, extracting, executing code snippets, and handling the outcomes of such executions.
- *Io module*: This module facilitates seamless data ingestion and result dissemination, thereby accelerating the iterative processes inherent in model training and evaluation.

- *Logger module*: This module plays a role in recording, tracking, and analyzing the usage, performance, and outcomes of API calls, making it an essential tool for monitoring, debugging, and optimizing application behavior.

2.9 Prompt Engineering

Prompt engineering is an evolving field focusing on crafting and refining prompts to effectively utilize LLMs for diverse tasks, including NLP (Louie Giray, 2023). White et al., (2023) Highlights that prompts act as a form of programming, enabling tailored outputs and interactions with LLMs. The paper by White et al., (2023) introduces a catalog of prompt engineering techniques to enhance interactions with LLMs. Key prompt techniques discussed by White et al., (2023) include:

- *Input Semantics*: Enhances how LLMs understand user inputs by using a structured or simplified language specific to a domain.
- *Output Customization*: Tailors the LLM's outputs to fit specific formats or styles. Techniques like "Output Automator" generate automated responses, "Persona" imbues the model with a consistent character or tone, and "Visualization Generator" produces graphical outputs.
- *Error Identification*: Helps pinpoint errors in the LLM's outputs. Techniques like "Fact Check List" generate a list of facts for verification, and "Reflection" prompts the model to reconsider and correct errors.
- *Prompt Improvement*: Enhances interaction quality by refining prompts. Techniques like "Question Refinement" clarify ambiguous queries, "Alternative Approaches" suggest different methods to tackle problems, "Cognitive Verifier" checks logical coherence, and "Refusal Breaker" encourages the model to overcome initial limitations.
- *Interaction*: Enriches user interactions with LLMs, making them more engaging. Techniques like "Flipped Interaction" involve the model posing questions to the user, "Game Play" integrates game elements, and "Infinite Generation" encourages the model to continually produce content.
- *Context Control*: Allows users to set specific contexts for the LLM to ensure relevance and accuracy. Techniques like "Context Manager" specify operational boundaries, such as setting a temporal or geographical limit on the data considered. For example, focusing on financial trends in Europe over the last five years.

Prompt engineering is a critical aspect of utilizing LLMs effectively. However, the order in which these samples are presented can significantly impact performance, sometimes making the difference between state-of-the-art results and random guessing (Lu et al., 2022). The inclusion of multi-agents architectures necessitates the development of more complex prompts in order to facilitate consistent operation and effective communication (Wang et al., 2024). Within the domain of integrating reasoning and acting in language models, Yao et al. (2023) introduce a prompting method called ReAct. This approach systematically orders prompting by allowing language models to generate both verbal reasoning traces and actions simultaneously. ReAct translates to the following steps (Yao et al., 2023):

- *Initial Prompt Setup*: The model receives a task, which could be a question or an interactive challenge, along with any relevant context or initial observations. This sets the stage for the task-solving process.
- *First Thought (Reasoning Trace)*: The model generates a "thought" or reasoning trace. This is an internal dialogue or reasoning step that the model uses to think through the task. It helps the model formulate a plan or hypothesis about what needs to be done.
- *First Action*: Based on the reasoning trace, the model proposes an action. This could be a physical action in an interactive environment or a query in a data retrieval task.
- *Observation from Action*: Once the action is taken, the model observes the outcome. This observation feeds back into the model's understanding of the situation.
- *Subsequent Reasoning*: With the new information obtained from the action's outcome, the model updates its reasoning. It might adjust its initial hypothesis or plan based on what it has learned. This step is crucial as it reflects a dynamic adjustment to the strategy, akin to human problem-solving where one adapts their approach based on new information.
- *Further Actions and Observations*: The cycle of reasoning and action continues, with each step informed by the previous one. The model iteratively generates new thoughts and actions until the task is either completed or a satisfactory solution is reached. Each action is accompanied by an observation that influences the next step in the reasoning process.
- *Completion of the Task*: The process culminates when the model successfully completes the task or reaches a logical stopping point. This could be determining the

correct answer to a question, successfully navigating to a location in an interactive environment, or realizing that the available information is insufficient to proceed further.

2.10 Conceptual Model

Based on the theoretical overview, several knowledge gaps have been identified. Previous research extensively covers earnings reports and sentiment analysis. However, it lacks in addressing the interplay between company reports, sentiment analysis around their release, and short-term stock trends. Figure 4 illustrates the relationships of the variables in the three research questions employed for this study. This study aims to investigate these relationships, specifically:

- *The relationship* between quarterly financial metrics and short-term market trends (1).
- *The relationship* between multi-agent sentiment analysis of news headlines and tweets and short-term market trends (2).
- *The relationship* between quarterly earnings metrics and multi-agent sentiment analysis (3).

Utilizing a multi-agent conversation for sentiment analysis, the study explores how different types of data relate to market behaviors.

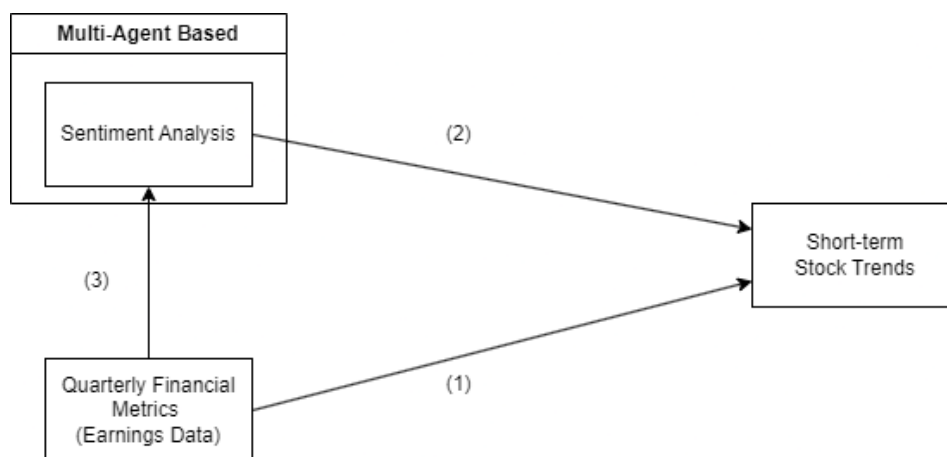


Figure 4: Conceptual Model

The relationships between the variables will be explored, utilizing correlation matrices, emphasizing the relationships of the variables during the 2022-2023 period. This approach seeks to provide an understanding of how financial earnings metrics and multi-agent-based

market sentiment relate to stock price movements, as well as how the release of quarterly financial earnings metrics relates to market sentiment.

2.11 Summary

This chapter provides an overview for understanding the combination of financial markets, investment theories, and AI. The chapter explores the applications of AI in sentiment analysis and the multi-agent conversation framework. The focus is on the use of the AutoGen library and prompt engineering for improving multi-agent driven sentiment analyses.

EMH posits that stock prices reflect all available information, setting a baseline for market efficiency studies. Behavioral finance introduces psychological factors that influence investor decisions, challenging the assumptions of EMH. This contrast highlights the potential for irrational behaviors and biases to create market inefficiencies. Advancements in AI, particularly NLP, enhance the understanding of human language in financial contexts. Sentiment analysis, a branch of NLP, decodes public perceptions from various data sources, including social media and news. This analysis allows researchers to categorize sentiment at document, sentence, and entity levels, providing a granular understanding of public sentiment. The review also introduces the multi-agent conversations, which utilizes collaborative intelligence among AI agents to improve decision-making processes in complex environments. This framework, facilitated by advancements like the AutoGen library, supports the dynamic interaction of AI agents in analyzing and responding to financial data. Prompt engineering emerges as a critical technique to optimize AI responses in these multi-agent systems, enabling more accurate and context-aware analyses of financial sentiment. The integration of these AI technologies and financial theories leads to the identification of gaps in current research, particularly in linking quantitative financial data and qualitative sentiment analysis to market trends, and financial data to market sentiment. These insights form the basis for the proposed conceptual recommendations in the study. This model aims to provide insights into how financial disclosures and public sentiment analysed through advanced AI techniques correlate with market behavior, particularly in the short term. This study is positioned to address these relationships using correlation matrices and multi-agent frameworks to offer a deeper understanding of market dynamics.

3 Quarterly Financial Metrics, News, Tweets, and Price Data

In this chapter the data used for this research is described. The purpose of this chapter is to describe the data used, how the data was obtained, and modifications applied to the data. Primarily using Refinitiv Workspace, financial reports, company focussed news headlines and tweets, and price data are gathered, in the 2022-2023 period. The data includes quarterly metrics, news headlines, tweets, and price data in the timespan of the release of company earnings calls each quarter. This chapter starts with describing the data specifications. Following, the three separate supplementary data files and the methodology to obtain and modify these datapoints are described. Finally, the merged data set that is based on the supplementary data files is described. This final dataset is used for the statistical analysis, using a correlation matrix.

3.1 Data Specifications

To collect the sentiment related to and performances of the companies utilized in this study, the data platform Refinitiv Workspace is used. Refinitiv Workspace is a financial data and analytics platform designed for professionals in the finance industry. It provides access to a wide range of financial information, analytics tools, and market data. (London Stock Exchange Group, n.d.). I have collected quantitative earnings data, qualitative news headlines and tweets, and price data for the period 2022-2023. The data is collected quarterly, corresponding to the release periods of the earnings data. Appendix A1 & A2 describe specific details, collection, and the perceived value of the data. The data is categorized into three supplementary data files: news headlines and tweets, quantitative financial earnings data, and stock price data. Finally, the three types of data are merged into a final dataset. This final dataset enables the analyses of the relationships between sentiment, quarterly financial earnings, and stock market trends. Data from 9 companies listed in the S&P 500 are considered. These companies were chosen, due to the diverse range of sectors, and them being the leaders in their industries. The analysis of these companies aims to provide an analysis of various sectors, offering a balanced perspective on market dynamics and financial trends across different industries. Table 1 provides the details of the companies analysed in this study (London Stock Exchange Group, n.d.).

Table 1: Specifications of the companies considered.

Company	Ticker Symbol	Company Sector
Apple Inc	AAPL	Computers, Phones & Household Electronics
Amazon.com Inc	AMZN	Diversified Retail
Berkshire Hathaway Inc	BRKa	Consumer Goods Conglomerates
Alphabet Inc	GOOGL	Software & IT Services
Eli Lilly and Co	LLY	Pharmaceuticals
Meta Platforms Inc	META	Software & IT Services
Microsoft Corp	MSFT	Software & IT Services
Nvidia Corp	NVDA	Semiconductors & Semiconductor Equipment
Tesla Inc	TSLA	Automobiles & Auto Parts

3.1.1 Financial Earnings Metrics (Quantitative data)

The first supplementary data file, `quantitative_metrics.csv`, consists of the quarterly financial earnings data for each company considered in this study. This dataset is created by manually merging individual financial summary data files obtained into a single collective file. This collective file contains the quarterly financial metrics for each company considered in this study (see Appendix A4 for a screenshot of the file). This manual construction was necessary due to the unique format of the raw data files retrieved from Refinitiv Workspace. The data points in this file span from 31-03-2022 to 31-12-2023 and encompass a wide range of metrics that offer insights into the performance of each company on a quarterly basis. Each column in the quantitative dataset integrates various dimensions of data, structured as follows:

- *Company*: The name of the S&P 500 company.
- *Quarter Date*: The date marking the end of a financial quarter.
- *Financial Metrics*: Includes detailed metrics from the company's quarterly financial summaries such as Revenue, EBITDA, Net Income, and other critical financial indicators.

The table in Appendix A3 provides a description of each quantitative metric included in the `quantitative_metrics.csv` file. Each metric is explained in terms of its definition, and its relevance. It is important to note that not all earning metrics from the raw data files obtained are included in this dataset. Data points are excluded due to duplicates and missing values

across the companies. This is done to ensure the integrity and accuracy of the analysis. These metrics are aimed to provide a detailed view of each company's financial performance and health, forming the basis for analyzing the relationship of quantitative metrics with short-term stock market trends, and market sentiment. This data considers the quarterly financial data from the period 31-03-2022 up until 31-12-2023. It is important to note that the companies included in this study have different quarterly earnings release dates. However, for consistency, the dates 31-03-202x, 30-06-202x, 30-09-202x, are used as the deadlines for Q1, Q2, Q3, and Q4 of each respective year.

3.1.2 News headlines and Tweets (Qualitative Data)

The datasets titled News_<company>_<year>.csv consist of .csv files that compile raw news headlines and tweets directly collected from the news monitor feature of Refinitiv Workspace, organized by each company. Refinitiv Workspace offers access to news data spanning 41 third-party news sources (London Stock Exchange Group, n.d.). The keywords used in Refinitiv Workspace are the company names, with the filter set to 'English' to retrieve data solely in English. The columns in the Qualitative data files include the following data points:

- *Date*: The date column contains the short date notation of publication of the news headlines and tweets. For example, '02/04/2022'.
- *News*: This column contains the content of the news headlines or tweet.

Each file in this dataset follows the naming convention News_Company_YYYY.csv, where <Company> represents the company name and <YYYY> represents the year. This results in two files per company, one for the year 2022 and one for 2023. The news headlines and tweets are collected on three specific days: 7 days before (Day -7), on the day of (Day 0), and 7 days after (Day +7) the release of quarterly company earnings report. This chosen timeframe aims to cover both immediate and subsequent market responses, essential for analysing the impact of sentiment on stock trends. A total of 19,715 news headlines and tweets are collected for multi-agent sentiment analysis. Table 2 describes the number of news headlines and tweets per company.

Table 2: News Headlines and Tweets per company

Company	News Headlines and Tweets
Apple Inc	3878
Amazon.com Inc	3255
Berkshire Hathaway Inc	752
Alphabet Inc	1412
Eli Lilly and Co	1110
Meta Platforms Inc	2530
Microsoft Corp	3285
Nvidia Corp	944
Tesla Inc	2549
Total	19715

Appendix A5 contains a snapshot of the ‘News_Apple_2022.csv’ file, the snapshot shows the columns included in each file. These news headlines and tweets, collected from Refinitiv Workspace, will be used to analyse the sentiment of news headlines and tweets with the multi-agent conversation framework. The average sentiment on the three specific days will be integrated into the merged final data file. It is important to note that Refinitiv Workspace did not include tweets for the 2022 period, this period only contains news headlines.

3.1.3 Price data

The final raw supplementary data file collected from Refinitiv Workspace includes price data of each company considered. The price data files will be utilized to determine short-term price trends. The dataset consists of a series of .csv files, each corresponding to stock price data for the 9 S&P 500 companies, covering the period from March 31, 2022, to December 31, 2023. The files are segmented quarterly and named according to the pattern Prices_[COMPANY].csv, where [COMPANY] is replaced by the company's stock ticker symbol, such as AAPL for Apple Inc. The columns of the price data files are provided in Table 3.

Table 3: Description of the Price Data File Columns

Column	Description
Quarter	The date marking the end of a financial quarter.
Date	The trading day for which the data is recorded (DD-MM-YYYY).
Close	The closing price of the stock on the given trading day.
Open	The opening price of the stock on the given trading day.
High	The highest price of the stock on the given trading day.
Low	The lowest price of the stock on the given trading day.
Volume	The number of shares traded on the given trading day.
3_day_MA	The 3-day moving average of the closing price, calculated using Python's Pandas library.
5_day_MA	The 5-day moving average of the closing price, calculated using Python's Pandas library.
7_day_MA	The 7-day moving average of the closing price, calculated using Python's Pandas library.

The price data covers the period around the quarterly financial statements, spanning from 7 days before the release to 7 days after the release (15 days in total) of quarterly company earnings metrics. This timeframe is chosen to capture both the immediate market reactions to the financial disclosures and the subsequent effects reflected in the news headlines and tweets. The calculation of the 3-day, 5-day, and 7-day moving averages are aimed to provide a smoothed view of price trends, reducing daily volatility. Moving averages are used to detect the direction of a trend and smooth out the price, volume fluctuations that can lead to possible confusion of trend interpretation (Chang et al., 2012). These moving averages, along with the closing prices, are used to determine the overall stock trends.

3.1.4 Data Collection, Cleaning, And Homogenization

As previously mentioned, data is solely extracted from Refinitiv Workspace, including financial summaries, price data, news headlines, and tweets. The quantitative and qualitative data are exported directly into Excel (.xlsx) files and transformed into .csv files. After extracting all the necessary data for analysis, the following steps are conducted:

- *Data Homogenization:* By standardizing the data, I eliminate variations caused by differing formats, scales, and units. This script was executed in Visual Studio Code (VSCoDe), utilizing the Python Pandas library for data manipulation. The code snippets

for homogenization used are provided in Appendix B1, containing an example for the company Apple.

- *Data Cleaning*: A Python script was developed to remove duplicates and empty values in the news headlines and tweets data. This step is performed to ensure the uniqueness of each entry. An example code snippet used for cleaning the data are provided in Appendix B2, for the company Apple.
- *Data Calculations*: The moving average calculation helped in indicating the price trend associated with each quarter. Following the moving average calculations, price trends are determined. This script was executed in Visual Studio Code (VSCode), utilizing the Python library pandas for data manipulation. Key snippets for the python code used are provided in Appendix B3 for the company Apple.

3.1.5 Code and Tools Used

The following tools and programming environments are utilized for data collection and modifications:

- *Refinitiv Workspace*: For data extraction.
- *.xlsx and .csv files*: For initial data storage and manual operations.
- *Visual Studio Code* (Microsoft Corporation, n.d.): Used as the integrated development environment for running Python scripts.
- *Python* (Python Software Foundation, 2023): The primary programming language used for scripting data processing tasks.
- *The Python scripts*: used for homogenizing data, the scripts include code that perform data homogenization, cleaning, and integration. Pandas Library (McKinney, 2010) operations are used to conduct cleaning and merging the data.

To ensure the reproducibility of the experiment, detailed comments were included within the code scripts to explain each operation and its purpose. The complete Python scripts along execution instructions are provided in Appendix B1, B2, and B3.

3.1.6 Merged Dataset

After collecting and modifying all the supplementary datafiles, I have merged all files into one .csv file. This final dataset serves as a central repository for analyzed, modified, and

standardized metrics. The file is named "Final_dataset.csv", a snapshot of the final dataset is provided in Appendix A6. The final dataset is structured with columns that integrate the following various dimensions of data:

- *Company*: The name of the S&P 500 company.
- *Quarter Date*: The date marking the end of a financial quarter.
- *Revenue, EBITDA, Net Income, etc.*: Financial earnings metrics from the company's quarterly financial summaries, see Appendix A3 for all included financial earnings metrics.
- *Average Sentiment Day -7, Average Sentiment on Release Day 0, Average Sentiment Day 7*: The average sentiment score, of the news headlines and tweets 7 days before, on the day of, and 7 days after the release of financial summaries are summarized per quarter in this column. These data points contain 3 columns of data per LLM utilized, which translates to 6 in total. Section 4.3 discusses the multi-agent sentiment analysis in detail.
- *Price Trend*: The trend in stock price over the quarter, categorized as 1 (up), -1 (down), or '0' (stable). The price trend is calculated with the closing price and the 3,5 and 7-day moving average. This is covered in detail in Section 4.4 "Price Trend Definition".

This final dataset contains 8 datapoints per considered company for all quarters, resulting in a final dataset of 72 rows. This final dataset serves as the basis for correlation analysis, to uncover the relationships of the quarterly financial metrics with short-term stock trends, multi-agent sentiment analysis with short-term stock trends, and the relationship of quarterly financial metrics and multi-agent sentiment.

4 Experimental Research Design

This chapter outlines the methodology used in this research, detailing the experimental research design employed. This section describes the materials, experiments, and procedures used for analyzing the relationships of the considered variables in this study. The chapter is structured to provide an overview of the research process, including the definition of the agents, setting-up of the multi-agent conversation framework, the method of sentiment analysis, and the statistical evaluation using a correlation matrix. This chapter first discusses the experimental

design. Following, each task of the agents conducting the sentiment analysis is described (definition of agents), working together for a collective sentiment analysis. After this, the multi-agent conversation is explained in detail. The chapter ends with a description of determining the price trends and the statistical analysis method employed.

4.1 Experimental Design

This study utilized an experimental research design to investigate the relationships of quarterly financial earnings, multi-agent sentiment, and short-term stock trends. Employing a mixed-methods approach, quantitative and qualitative research was conducted. Quantitative research is exemplified by social surveys and experimental investigations. It focuses on numerical data, statistical analysis, and seeks to test hypotheses derived from theoretical frameworks (Bryman, 2003). In this study, I collected quantitative financial metrics for 9 S&P 500 companies on a quarterly basis, details of the companies are presented in Table 1. Qualitative research, in contrast, emphasizes understanding the complexity of social phenomena through detailed,

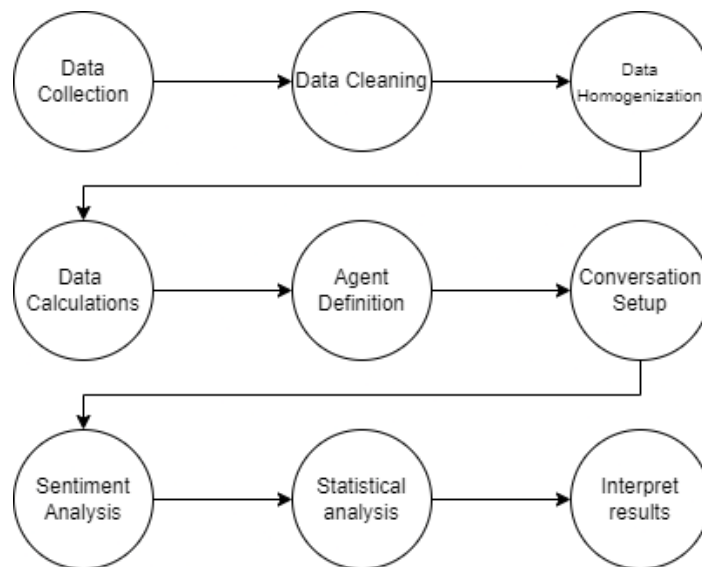


Figure 5: Experimental Research Design Steps

contextual, and interpretive analysis (Bryman, 2003). This study incorporated qualitative analysis on text data from news headlines and tweets, by performing multi-agent sentiment analysis. The sentiment analysis was conducted using a multi-agent conversation framework, which interpreted the sentiment based on specific sentiment words, sarcasm, and emotions expressed in the textual data. Figure 5 illustrates the experimental research design steps, which are detailed in the following sections. The data collection, cleaning, homogenization, and calculations are discussed in detail in Chapter 3. By combining quantitative financial metrics

and qualitative sentiment analysis, this study aimed to leverage the strengths of both approaches to provide a understanding of the relationships between the variables employed in this study. The research steps were facilitated using various tools. Namely, Refinitiv Workspace, Python 3.10.12 (Python Software Foundation, 2023), the Pandas library (McKinney, 2010), and the AutoGen Library (Microsoft, 2024) in the IDE VSCode (Microsoft Corporation, n.d.). The multi-agent conversation framework utilized two different LLMs as key components. This resulted in the news headlines and tweets being analyzed twice, each with a different LLM as key component. One LLM was accessed through Groq Cloud (Groqcloud. Groq Cloud, n.d.), and one was accessed using the Open AI API (OpenAI, n.d.-b). To answer the research questions of this study, the approach involved:

- Implementing a multi-agent conversation framework for multi-agent sentiment analysis
- Performing sentiment analysis within the specified timeframe around the release of quarterly financial reports.
- Using a correlation matrix to explore relationships between financial metrics, multi-agent sentiment, and short-term stock price trends.

These facilitated exploratory research steps aimed to uncover the relationships between quantitative quarterly metrics, sentiment analysis, and stock market trends. The goal of this study is to contribute to the broader understanding of these interactions while using innovative tools for conducting sentiment analysis. The correlation matrix was employed as the primary tool to investigate these relationships.

4.2 Defining the Agents

For conducting sentiment analysis on the corpus of news headlines and tweets, I defined various agents that participated in a sentiment analysis team. The aim of the agent roles was to divide the sentiment analysis into small and specific tasks, leveraging the specialized expertise of each agent. The primary objective was to determine the sentiment of the news headlines and tweets related to the companies considered in this study, specifically around the release of quarterly company earnings release. Each agent executed its task, contributing to the final sentiment analysis. The agent roles are defined as follows:

- *Words-analyst*: The words analyst is an agent specialized in determining the specific sentiment of words in news headlines and tweets. Its role is to analyze the sentiment conveyed by each word in the provided news headlines or tweets.
- *Sarcasm analyst*: The sarcasm analyst specializes in irony and sarcasm. It analyzes the news headlines or tweets and detects sarcasm in the news headline or tweet provided. For example, "Although 'love' is generally positive, the phrase 'love to hate' suggests a negative sentiment."
- *Emotion Analyst*: The emotion analyst goes beyond positive or negative, this agent identifies specific emotions like joy, anger, fear surprise etc., which are detected by analyzing the text.
- *Sentiment analyst*: The sentiment analyst is an assistant agent responsible for scoring the sentiment of news headlines and tweets based on three aspects: the sentiment of the words, the presence of sarcasm or irony, and the emotions conveyed. The sentiment analyst incorporates feedback from the sentiment critic to revise the sentiment scores. If no feedback is received, the initial scores remain unchanged.
- *Sentiment Critic*: The sentiment critic provides critique and improvement points based on the sentiment scores assigned by the sentiment analyst. For example: "I agree with most of the scores assigned to each news headline and tweet, however I suggest that the second headline is mostly positive instead of neutral." If the sentiment critic agrees with the sentiment analysis done, it will output: 'All good!'.

These agents collaborated in the conversation workflow, conducting a collective sentiment analysis. The Words Analyst determined the sentiment of words, laying the groundwork for more complex analyses. As stated by Liu, (2012), the most important indicators are sentiment words, also called opinion words. The Sarcasm Analyst adjusted for sarcasm and ensured that sentiment scores reflect the true intent behind the words. Sarcasm is often labeled as a sentiment shifter, and as one of the challenging tasks for LLMs (Liu, 2012; Zhang., 2023). However, with combining the sentiment of words and sarcasm I aimed to solve this challenge through the output of the words-analyst and the sarcasm analyst. The Words-Analyst focuses on individual word sentiment. The Sarcasm Analyst then identified the sarcastic tones with the knowing the sentiment of the specific words in the news headline or tweet. The Emotion Analyst captured a range of emotions such as joy, anger, fear, and surprise. Closely related to sentiment and

opinions lies emotion (Liu, 2012), this agent covers this aspect of sentiment analysis. The Sentiment Analyst combined the insights from the Words Analyst, Sarcasm Analyst, and Emotion Analyst. This approach considered various aspects of sentiment to produce sentiment scores, considering the analysis of the previous agents. The Sentiment Critic reviewed and critiqued the sentiment scores assigned by the Sentiment Analyst. It suggested improvements to address potential errors or oversights. Feedback mechanisms are proven to be performance enhancers, and part of prompting strategies in previous research (Sun et al., 2023; White et al., 2023).

4.3 Multi-Agent Sentiment Analysis: Conversation Set-up

Utilizing the AutoGen library (Microsoft, 2024), tools for deploying multi-agent systems through computer code were accessed. The AutoGen library enabled programming conversation patterns for the multi-agent sentiment analysis. Agents use LLMs as key components to understand and react to messages (Microsoft, 2024). As stated previously I employed two LLMs for the agents to utilize for analyzing sentiment in news headlines and tweets. The two different LLMs were employed to detect any differences in sentiment analyzed per LLM. Specifically, Meta's LLaMa3 8B Instruct, accessed through Groq Cloud, and OpenAI's GPT-3.5, accessed via the OpenAI API, were used (Groqcloud. Groq Cloud, n.d.; HuggingFace, n.d.; OpenAI, n.d.-b). The cloud-based models were utilized to enhance output quality and processing speed. Due to limits in the context window of both LLMs, the 19715 news headlines and tweets are processed in batches. Testing the multi-agent framework revealed a limit in the number of news headlines and tweets that could be effectively analyzed in one go. Table 4 describes the batch size, context window, and knowledge cutoff for each LLM.

Table 4: Sentiment analysis batch size

Model	Batch Size	Context Window	Approx. Words	Knowledge cutoff
gpt-3.5-turbo-0125	5	16385 tokens	~12,500 words	September, 2021
Llama-3-8B-Instruct	5	8000 tokens	~6,000 words	March, 2023

Testing showed that exceeding a batch size of 5 causes issues. For example, the sentiment critic rarely provided feedback, and the sentiment analyst did not consider the analyses of other agents for scoring. Increasing the batch size resulted in agents excluding certain news headlines

and tweets. For instance, with a batch size of 25, agents only considered the first 10 news headlines and tweets of the batch. The code, utilizing the AutoGen library, is presented in Appendix B5.

4.3.1 Group Chat Workflow

The sentiment analysis with the multi-agent system, employed a predefined workflow. This workflow facilitated agent collaboration, with the agents using the outputs of one another to conduct sentiment analysis through a group chat pattern. AutoGen allowed for general conversation patterns, the pattern used in this study is called ‘group chat’. The ‘group chat’ in the AutoGen library allowed for all agents specified to contribute to a single conversation thread and contribute to the general task at hand. Simply, the ‘group chat’ functionality allows for collaboration between agents. This feature is useful for integrating diverse inputs and analyses from various agents into a collective output (Microsoft, 2024). Figure 6 illustrates the

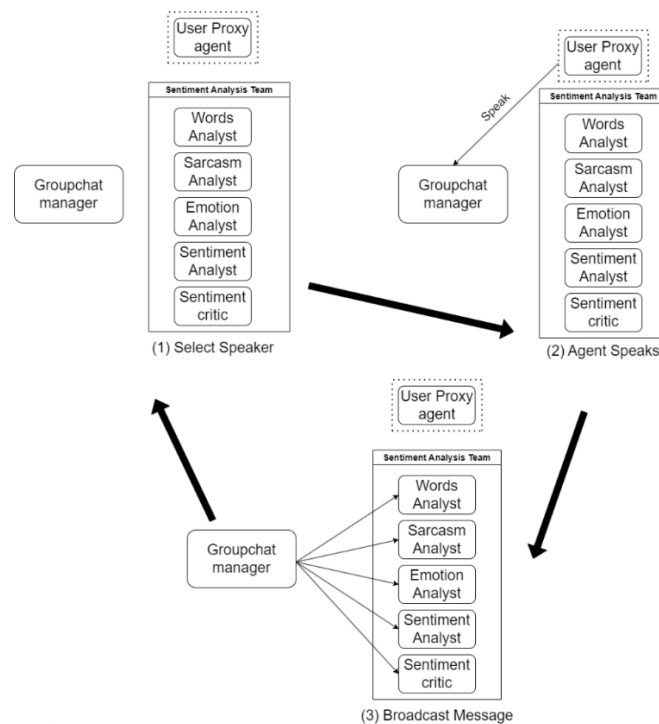


Figure 6: Group Chat Workflow, based on the AutoGen Documentation (Microsoft, 2024)

workflow of the group chat for the collaborative sentiment analysis conducted by the agents of this study. I created this figure based on the illustrative example of the group chat in the AutoGen documentation (Microsoft, 2024).

For sentiment analysis, a special agent type called the ‘GroupChat Manager’ orchestrated the conversation pattern and selected speakers accordingly. The GroupChat Manager initiates the workflow by first selecting the User Proxy Agent. The User Proxy Agent served as an

administrative agent, providing the human prompt to the GroupChat Manager to broadcast the human prompt to the sentiment analysis team, the human prompt is discussed in section 4.3.2. After broadcasting the human prompt, the sentiment analysis team executed their assigned tasks and collaborated to provide the sentiment scores. The workflow of the group chat followed these steps:

1. *Select Speaker*: The GroupChat Manager selects an agent to speak.
2. *Agent Speaks*: The Agent Provides its output to the group chat manager.
3. *Broadcast Message*: The GroupChat Manager broadcasts the message to all other agents in the group chat. The process repeats until the conversation is finished.

The GroupChat Manager selected agents in a ‘round-robin’ fashion, following a predefined order. This ensured that each agent had an opportunity to contribute to the conversation in a systematic manner. This approach allowed each agent to perform its specialized task sequentially, contributing to the sentiment analysis. Building on the ReAct style behavior of Yao et al. (2022), I propose a structured multi-agent sentiment analysis method. Agents considered the output of other agents, provide reasoning for their thoughts, and performed specific actions. The sentiment critic differed slightly, as its role was to observe the sentiment analysis and provide critique or feedback. Figure 7 illustrates the steps of the multi-agent sentiment analysis. The GroupChat Manager broadcasted each agent's output to the other agents as explained previously. The followed steps of the multi agent sentiment analysis were conducted as follows:

0. The user proxy provided the initial prompt compiled by the human.
1. The words analyst determined sentiment of specific words in the news headline or tweet.
2. Following, the sarcasm analyst detected sarcasm in the news headline or tweet. If there was no sarcasm found, the agent returned ‘No Sarcasm Detected.’
3. The Emotion analyst conducted emotion analysis on the news headline and tweet. Which entailed emotions such as ‘Joy’, ‘Optimism’ etc. If there was no emotion detected, the agent returned ‘No Emotion Detected.’
4. The Sentiment analyst provided a sentiment score based on the analysis conducted by the Words Analyst, Sarcasm Analyst, and the Emotion Analyst.
5. The sentiment critic observed the sentiment analysis done and provided critique and feedback on the sentiment score provided by the sentiment analyst (5a). If there was an agreement on the sentiment analysis done, the critic outputted ‘All good!’, this keyword terminated the conversation (5b).
6. If the sentiment analyst received feedback, it will incorporate the feedback and assign a new score to the news headlines and tweets. This completes the sentiment process.

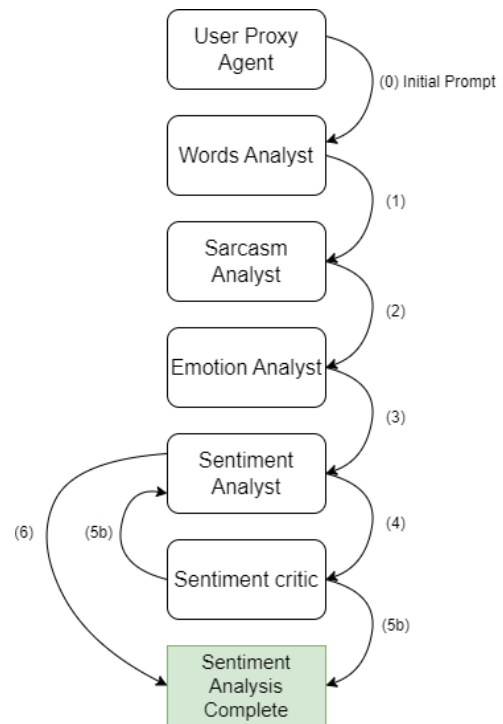


Figure 7: Sentiment Analysis Steps Executed.

Appendixes D1 & D2 contain example conversations of the agents conducting sentiment analysis for each LLM considered in this study. The examples demonstrate the iterative nature of the sentiment analysis process and how feedback was integrated to refine the final sentiment scores. The appendix section includes detailed examples of how each LLM processed input and collaborated with other agents.

4.3.2 Tailoring the Prompts (Human Prompt)

The human prompt, provided by the User Proxy Agent as detailed in section 4.3.1, was designed to guide the agents to follow a specific pattern. The human prompt served as a

collective task directive broadcasted to each agent. This prompt instructed the agents to conduct collective sentiment analysis on news headlines and tweets related to the companies under consideration. The human prompt can be broken down in the following parts:

- *“Work together to conduct sentiment analysis on the provided news headlines and tweets relating to the company <Company name is provided here>”*: This part of the prompt acted as the specific task assigned to the sentiment analysis team. The prompting Technique applied is the ‘Initial Prompt Setup’ (Yao et al., 2023). This prompt also encouraged Interaction (White et al., 2023).
- *“There are specific tasks assigned to each of you, you try to complete this task at the best of your capabilities”*: This part of the prompt assigned a ‘Persona’ (White et al., 2023) to each of the agents. This persona is specified in the system message of the agents.
- *“Note: the goal of the sentiment analysis is to discover the sentiment of the news headline and tweet towards the company”*: This part of the prompt specified the goal of the sentiment analysis, ‘Managing the Context’ (White et al., 2023) of the sentiment analysis team.
- *“The following format should be applied: Always include the provided news headlines and tweets, Always provide reasoning for your thoughts”*: This part of the prompt ‘Customized the Output’ (White et al., 2023). It specified the output to follow for each agent.
- *“For the sentiment analysis, you use the following rating scales: -1 = Negative -0.5 = Mostly Negative 0 = Neutral 0.5 = Mostly Positive 1 = Positive”*: This part of the prompt specified the sentiment rating scales to follow. Customizing the Output of each agent (White et al., 2023), it aimed to correctly assign a standardized sentiment score to each news headline and tweet.

I compiled this human prompt to provide a collective context to the sentiment analysis team. The prompt was mainly designed to tailor the outputs of the agents and specify the goal of the task at hand, specifically collaborative sentiment analysis. By motivating the agents to work collaboratively, it aimed to leverage the specific strengths and actions of the agent.

4.3.3 Tailoring the Prompts (Agent Messages)

The agents utilized in this study were prompt engineered. Simply put, the LLM based agents were programmed through prompts. Each agent can be programmed via a system message to tailor its output (Microsoft, 2024). Variations in prompt design can substantially influence the performance of LLMs (Zhang et al., 2023a). With this in mind, I constructed a consistent set of prompts for both LLMs and LLM-based agents to conduct sentiment analysis. The steps for constructing the prompts were based on the prompt improvement catalog & ReAct (White et al., 2023; Yao et al., 2023). More specifically, each agent was motivated to provide reasoning for its analysis. Additionally, the process for constructing the prompts was done with continuously improving the prompts. Figure 8 illustrates the prompt design methodology employed.

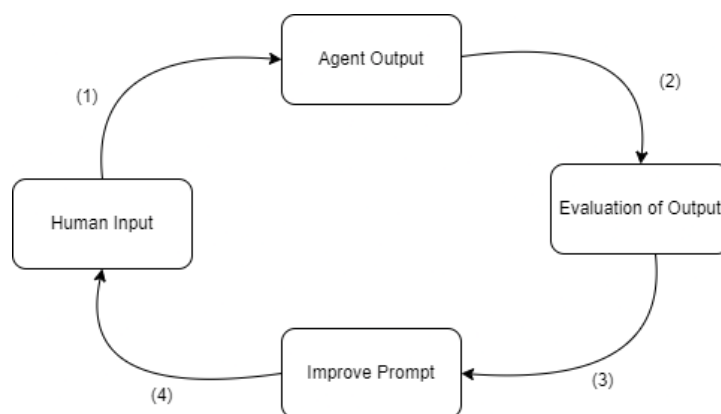


Figure 8: Steps Conducted for Tailoring the Agent Specific Prompts

The process began by defining simple tasks for each agent using a human prompt, as illustrated (1). For instance, the initial prompt for the sarcasm analyst was: “*You are the sarcasm analyst; you specialize in finding sarcasm and irony in news headlines and tweets.*” After running the first test and evaluating the results (2), the prompts were redesigned and adjusted to correct undesirable actions by the agents (3). During the evaluation of the sarcasm analyst, it was observed that the agent was transforming news headlines and tweets into sarcastic versions. To address this issue, the prompt was revised to include: “*You do not transform the news headline into sarcasm; you only detect sarcasm.*” This adjustment (4) ensured that the agent identified sarcasm without altering the original content. The revised prompt was then re-inputted (1), completing the cycle and refining the agent's performance. I repeated this iterative process for

each agent until the desired behavior for each agent was achieved. The transformation of the prompts for each agent are specified in Appendixes C1 up until C5.

4.4 Sentiment Analysis

The sentiment analysis in this study used an Aspect-Based Sentiment Analysis, it focused on sentiment towards specific entities or aspects rather than entire sentences or documents (Liu, 2012). This approach is utilized in this study as I analyze sentiment in news headlines and tweets about nine companies. The goal is to determine the sentiment towards each company in the news headlines and tweets. I calculate the aggregate sentiment polarity scores around each quarterly financial statement release for each company. This includes measurements at Day -7, Day 0, and Day 7.

- *Polarity* indicates the degree of opinion towards a topic. In this study, the topics the news headlines and tweets encompassed were the 9 S&P 500 companies.
- *Polarity scores* ranged from -1 (negative) to 1 (positive), with 0 representing a neutral sentiment. I included intermediate scores of -0.5 (mostly negative) and 0.5 (mostly positive) for a finer-grained analysis, a common scoring method for aspect-based sentiment analysis.

After conducting the sentiment analysis, the mean of the sentiment scores per day were calculated. This resulted in three data points per LLM utilized, corresponding to an average sentiment for each specific day the news headlines and tweets were collected on. The timeframe covered the previously specified periods around the release of the quarterly financial statements. It allowed for a nuanced sentiment analysis to uncover the relationship of market perceptions and their potential relationship on short-term stock trends. Appendixes D1 & D2 contain the full conversations between agents, illustrating the applied methodology and the use of multiple agents in the sentiment analysis process.

4.5 Price Trend Definition

With the aim to discover the relationship between quarterly financial metrics, sentiment analysis, and short-term stock trends, I calculated the short-term stock trends. The methodology employed aimed to identify the direction of the trend. I used the difference between the close price, 3-day moving average (MA), 5-day MA, and 7-day MA to define the price trend. The

price data ranged from 7 days before to 7 days after the financial summary release, covering a total of 15 days, excluding non-trading days. For each period, the initial and final values were calculated as follows:

- s_i : Closing price on day i .
- $MA3_i$: 3-day moving average on day i .
- $MA5_i$: 5-day moving average on day i .
- $MA7_i$: 7-day moving average on day i .

To assess the stock market trends, moving averages were employed. This smoothing effect clarified the overall market direction by reducing noise from daily price changes. Using moving averages helped in identifying consistent trends, aimed to make the trend analysis more reliable. Specifically, the following values were considered:

- s_{start} : Closing price at the start of the period.
- s_{end} : Closing price at the end of the period.
- $MA3_{start}$: First valid 3-day MA value in the period.
- $MA3_{end}$: Last 3-day MA value in the period.
- $MA5_{start}$: First valid 5-day MA value in the period.
- $MA5_{end}$: Last 5-day MA value in the period.
- $MA7_{start}$: First valid 7-day MA value in the period.
- $MA7_{end}$: Last 7-day MA value in the period.

To determine the up trend, the closing price and moving averages were examined on an increase from the beginning to the end of the fifteen-day period (Day -7 up until Day 7). This translates to the following formula:

$$up_{signals} = \sum (\{s_{end} > s_{start}\}, \{MA3_{end} > MA3_{start}\}, \{MA5_{end} > MA5_{start}\}, \{MA7_{end} > MA7_{start}\})$$

Down trends were determined by examining if the closing price and moving averages have decreased from the start to the end of the period, this is illustrated by the following formula:

$$down_{signals} = \sum (\{s_{end} < s_{start}\}, \{MA3_{end} < MA3_{start}\}, \{MA5_{end} < MA5_{start}\}, \{MA7_{end} < MA7_{start}\})$$

These formulas calculated the sum of the signals, indicating whether the stock is in an upward or downward trend. Each signal represented a comparison between the end and start values of the closing price and the moving averages. A positive comparison contributes to the uptrend signals (end value is bigger than starting value), while a negative comparison contributes to the downtrend signals (end value is smaller than starting value). After calculating the sum of the signals, the trend for the quarterly period is determined. At least three of the signals must indicate the same direction for it to be considered the trend. The trend T for the period is classified based on the number of signals as follows:

$$T = \begin{cases} 1, & \text{if } up_{signals} \geq 3 \\ -1, & \text{if } down_{signals} \geq 3 \\ 0, & \text{otherwise} \end{cases}$$

In summary, trend T is classified based on the majority of signals indicating an uptrend, downtrend, or stable trend. A threshold of 0.5% is applied to exclude small fluctuations (+0.5% for up trends and -0.5% for down trends). This threshold filters out noise and minor price movements, focusing the analysis on significant changes. By excluding insignificant changes,

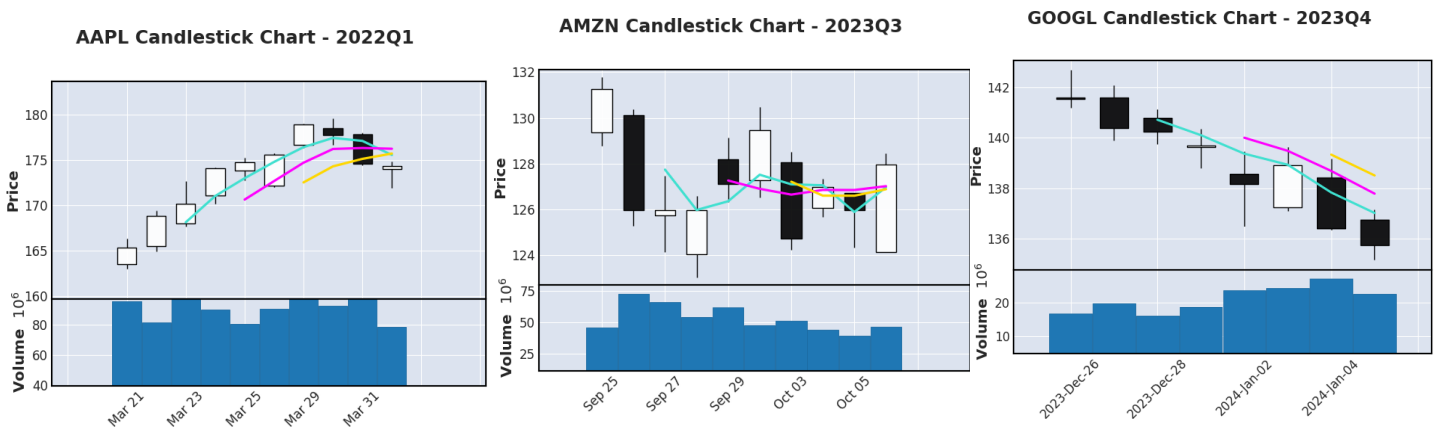


Figure 9: Quarterly Price Trends for the companies AAPL, AMZN, GOOGL (left to right: Up trend, Stable trend, Down trend)

the threshold aims to provide accurate trend classification. Figure 9 shows examples of price patterns classified as an uptrend, downtrend, and stable trend for the respective

This method used moving averages and closing prices over a specific period, offering a structured approach for trend analysis. Including a threshold helped to maintain consistency in

identifying significant trends, and to determine stable trends. The code snippets for calculating the price trends per quarter for each company are presented in Appendix B4.

4.6 Statistical Analysis

To discover the relationships between quarterly financial reports, sentiment analysis, and stock prices, a correlation matrix was employed. In its broadest definition, correlation measures the relationship between two variables. When two variables are correlated, a change in one variable is linked to a change in the other. This change can be positive or negative. This phenomenon is known as correlated data (Schober et al., 2018). The use of correlation matrices has a long history in finance, highlighted by Markowitz (1991), who uses them as a key concept in constructing optimal portfolios. In this study, a collective correlation matrix for all companies was created. The collective correlation matrix focused on the relationships of the variables across all companies in the dataset. The matrix quantified the relationships between variables, providing numerical values for each pair. By aggregating data from multiple companies, the collective matrix offered a broader view of how financial metrics and sentiment analysis relate to stock market trends. These values range from -1 to 1, where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no correlation. This analysis helped in identifying the relationships between financial metrics, sentiment scores, and stock price movements. The interpretation of the correlation coefficients can be guided by the approach described by Schober et al., (2018) as outlined in Table 5.

Table 5: Interpretation of Correlation Coefficients

Observed Correlation Coefficient	Interpretation
0.00-0.10	Negligible Correlation
0.10-0.39	Weak Correlation
0.40-0.69	Moderate Correlation
0.70-0.89	Strong Correlation
0.90-1.00	Very Strong Correlation

These guidelines helped in categorizing the strength of relationships in the collective correlation matrix. A coefficient between 0.00 and 0.10 indicates a negligible correlation. A coefficient between 0.10 and 0.39 indicates a weak correlation. A moderate correlation ranges from 0.40 to 0.69. Strong correlations fall between 0.70 and 0.89. Very strong correlations are

between 0.90 and 1.00. The correlation matrix in this study is created using the Seaborn and Matplotlib libraries. These libraries provide tools for visualizing data relationships. Heatmaps generated by these tools visually represent the correlation coefficients.

4.7 Declaration of Competing Interest

The author declares that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

5 Results

In this section, the results of the relationships between the variables of this study are presented, as derived from the correlation matrices. By examining the correlation matrices, I identified and analysed the strength and the direction of relationships between various quarterly financial metrics, multi-agent sentiment, and stock market trends. First, the correlations of quarterly financial metrics with price trends are presented. Following this, the chapter delves into the relationships between multi-agent sentiment analysis and price trends, considering the correlation coefficients from three distinct time points. This comparison analyses the insights from various LLMs utilized in the multi-agent sentiment analysis. The chapter ends with the relationships of specific quarterly financial earnings metrics and multi-agent sentiment.

5.1 Correlation Analysis

The scope of this research was to uncover the relationships among various variables. Specifically, the relationship between quarterly financial metrics and stock market trends, multi-agent sentiment analysis and stock market trends, and the relationship of quarterly financial metrics release on multi-agent sentiment analysis were examined. Figure 10 shows

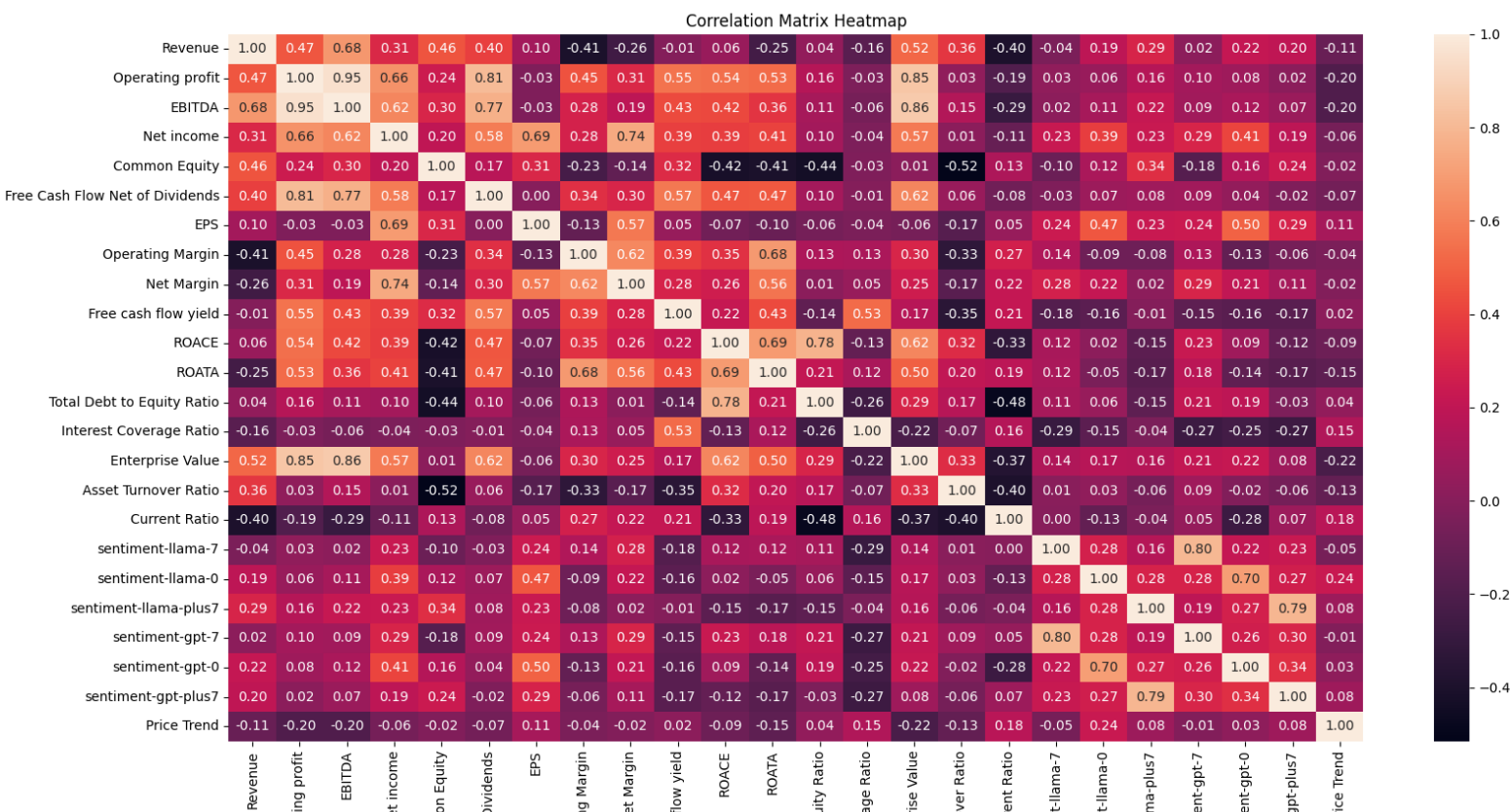


Figure 10: Collective Correlation Matrix of the Final Dataset

the compiled correlation matrix based on the final dataset. This matrix illustrated the relationships between financial metrics, multi-agent sentiment analysis using two different LLMs, and stock market trends. The correlation matrix can be interpreted as follows:

- *Positive correlations* appear lighter in colour, ranging from light purple (weak) to beige (strong).
- *Negligible Correlations* (Correlation < 0.10) appear in a purple colour.
- *Negative correlations* appear darker, ranging from darker purple (weak) to a black colour (strong).

The correlation matrix aimed to identify the strength and direction of the relationships (i.e., associations) between the variables. The correlation matrix illustrated the relationships of the variables on a collective basis. It considered all 72 datapoints of the companies Apple, Amazon, Berkshire Hathaway, Alphabet (Google), Eli Lilly, Meta, Microsoft, Nvidia, and Tesla.

5.2 Financial Metrics and Stock Trends

The correlations between financial metrics and short-term stock trends contained unique patterns for each metric, including both positive and negative correlations. Focusing on the metrics that exhibit significant correlations, Figure 11 provides the details of the specific financial metrics and their corresponding (non-negligible) coefficients with short-term stock trends.

The analysis revealed weak positive correlations for several financial metrics. Earnings Per Share (EPS) had a correlation coefficient of 0.11 with price trends. The Interest Coverage Ratio showed a correlation coefficient of 0.15 with price trends. The Current Ratio had a correlation coefficient of 0.18 with price trends. These weak positive correlations suggested that the increase of these specific financial metrics was associated with a modest rise of stock trends.

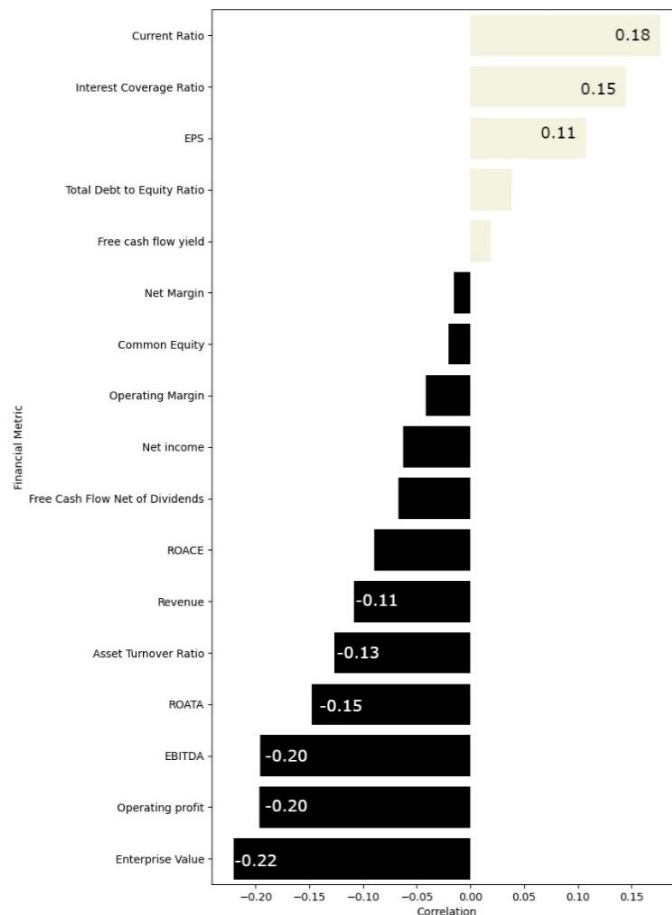


Figure 11: Correlations of the Financial Metrics and Short-Term Stock Trends

Several financial metrics displayed negative correlations with stock trends. Several financial metrics exhibited weak negative correlations with price trends. Revenue, Asset Turnover Ratio, and ROATA have correlation coefficients of -0.11, -0.13, and -0.15, respectively. EBITDA and Operating Profit each have a correlation coefficient of -0.20, while Enterprise Value shows a correlation of -0.22. These findings suggest that the relationship of these metrics indicated an inverse relationship with short-term stock trends.

5.3 Multi-Agent Sentiment Analysis and Stock Price Trends

While financial metrics offered quantitative insights into a company’s performance, sentiment analysis offered a qualitative perspective. By capturing market sentiment and investor attitudes,

the multi-agent sentiment analysis aimed to reveal how external perceptions and market psychology relate to short-term stock trends. Table 6 presents the correlation coefficients for each LLM on three specific days: seven days before the financial release (Day -7), on the day of the release (Day 0), and seven days after the release (Day 7). The table contains the correlation coefficients of multi-agent sentiment and short-term stock prices conducted by Llama 8B Instruct as key component and with GPT 3.5 as key component.

Table 6: Correlation Coefficients of Multi-Agent Sentiment and Short-term Stock Trends

Specific Day	Llama 3 8B Instruct	GPT 3.5
Day -7	-0.05	-0.01
Day 0	0.24	0.03
Day 7	0.08	0.08

While both LLMs showed negligible correlations on Day -7 and Day 7, sentiment analysis on the day of the financial release (Day 0) showed different results. Llama 8B Instruct had a weak positive correlation coefficient of 0.24, indicating that positive sentiment on the release day correlates with a modest increase in stock prices. In contrast, GPT 3.5 presented a negligible correlation of 0.03 on Day 0. These findings highlight that sentiment analysis can vary significantly depending on the timing relative to financial releases, particularly on the day of the release. It is crucial to note the variations in how each LLM utilized as a key component for the multi-agent sentiment analysis correlates with stock prices, particularly on the day of financial release. The difference in correlations on Day 0 between Llama 8B Instruct and GPT 3.5 emphasized the importance of considering both the timing and the specific model used in sentiment analysis, when assessing the association of sentiment with short-term stock trends.

5.4 Financial metrics and Multi-Agent Sentiment

The timeframe of the multi-agent sentiment analysis was conducted in conjunction with the quarterly earnings call release dates. Several significant correlations were observed between the financial metrics and sentiment recorded seven days before the release of quarterly financial results. Figure 12 illustrate the observed significant correlations of the financial metrics with the sentiment for both LLMs utilized as key component.

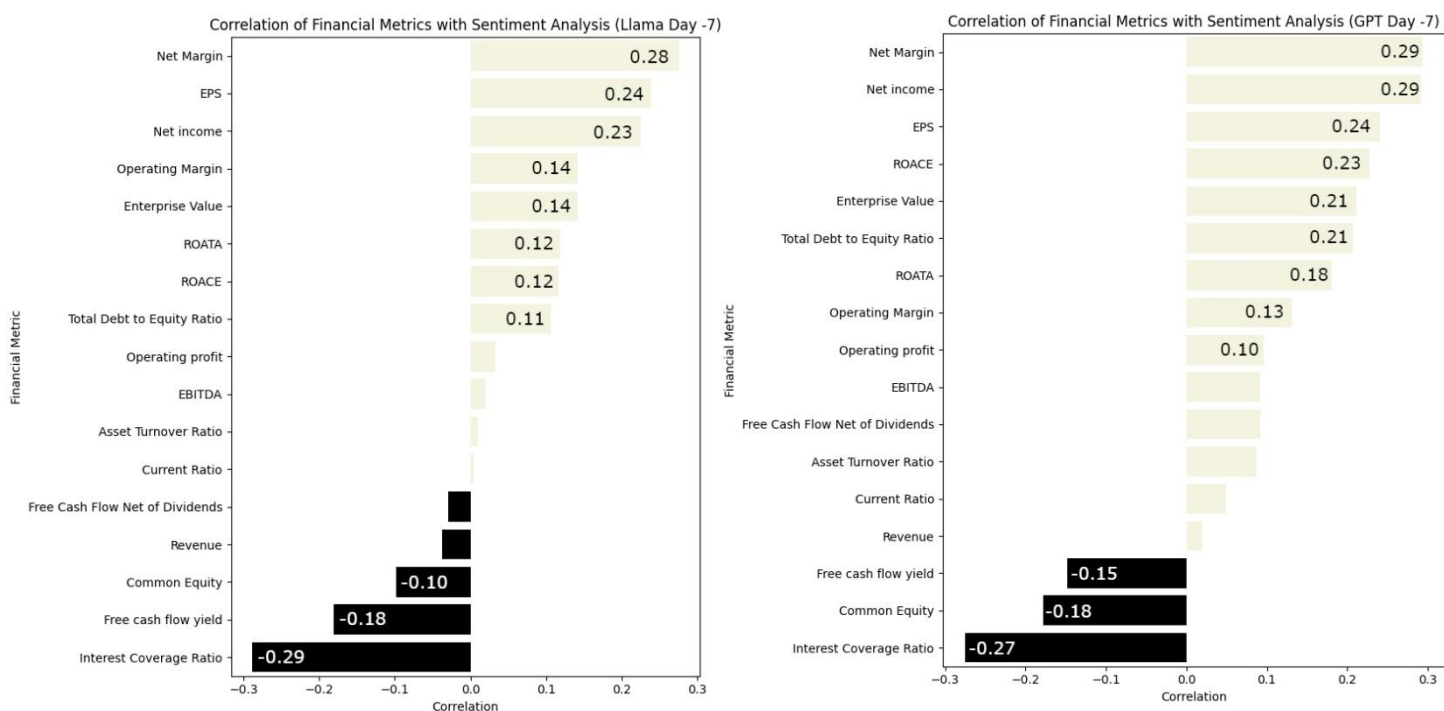


Figure 12: Significant correlations financial metrics and sentiment (Day -7) Llama 8B Instruct (left) sentiment and GPT 3.5 sentiment (right).

Both LLMs indicated the same directional relationships between financial metrics and multi-agent sentiment. Llama 8B Instruct showed notable positive correlation coefficients for Net Margin, EPS, and Net Income. These metrics were recorded seven days before the release of transcripts. These were weak positive correlations. This indicates that an increase in these financial metrics is associated to a modest positive sentiment. GPT-3.5 showed similar positive correlations for Net Margin, EPS, and Net Income. Additionally, GPT-3.5 showed weak positive correlations with ROACE, Enterprise Value, and Total Debt to Equity Ratio. Both LLMs displayed weak negative correlations with the Interest Coverage Ratio. The LLMs differed in their correlations with Common Equity and Free Cash Flow Yield. GPT-3.5 showed a stronger negative correlation with Common Equity. Llama 8B Instruct showed a stronger negative correlation with Free Cash Flow Yield. The weak negative correlation suggests an inverse relationship between the variables. To complement these insights, I explored the correlations on Day 0. I aimed to capture the direct impact of specific financial metrics on market sentiment. Figure 13 shows the observed correlations for both LLMs on Day 0.

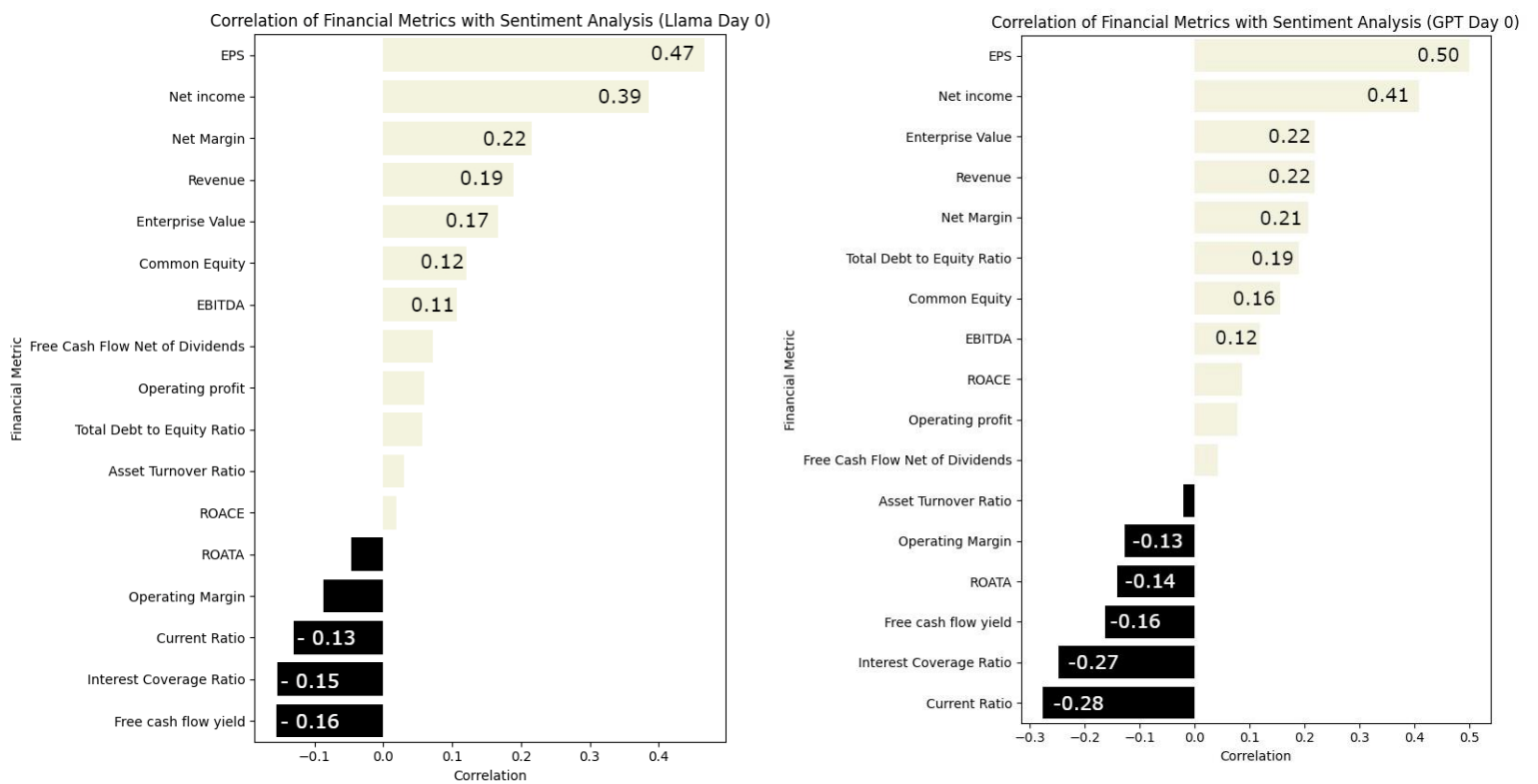


Figure 13: Significant correlations financial metrics and sentiment (Day 0) Llama 8B Instruct (left) sentiment and GPT 3.5 sentiment (right).

On the day of the financial release (Day 0), the correlation analysis for both Llama 8B Instruct and GPT-3.5 revealed notable patterns in the relationships between financial metrics and multi-agent sentiment. Both models showed moderate positive correlations with EPS and Net Income. Additionally, there was a symmetry in weak positive correlations for both models. Namely, with Net Margin, Revenue, Common Equity, EBITDA, and Enterprise Value. However, there is a difference: GPT-3.5 shows a weak positive correlation with Total Debt to Equity Ratio, while Llama 8B Instruct deems this metric to have a negligible correlation. These positive correlations indicated an association where higher values in these financial metrics are linked in symmetry to positive sentiment, and vice versa. Regarding negative correlations, both models exhibited weak negative correlations with Free Cash Flow Yield, Interest Coverage Ratio, and Current Ratio, varying in strengths per model. These weak negative correlations suggest a modest inverse relationship between these financial metrics and sentiment. The moderate correlations observed with EPS and Net Income suggest these metrics are more closely linked to sentiment changes on the day of the financial release. On Day +7, the patterns of correlation reveal additional insights into the longer-term impact of specific financial metrics

on market sentiment. Figure 14 illustrates the correlations of the financial metrics with multi-agent sentiment conducted seven days after the release.

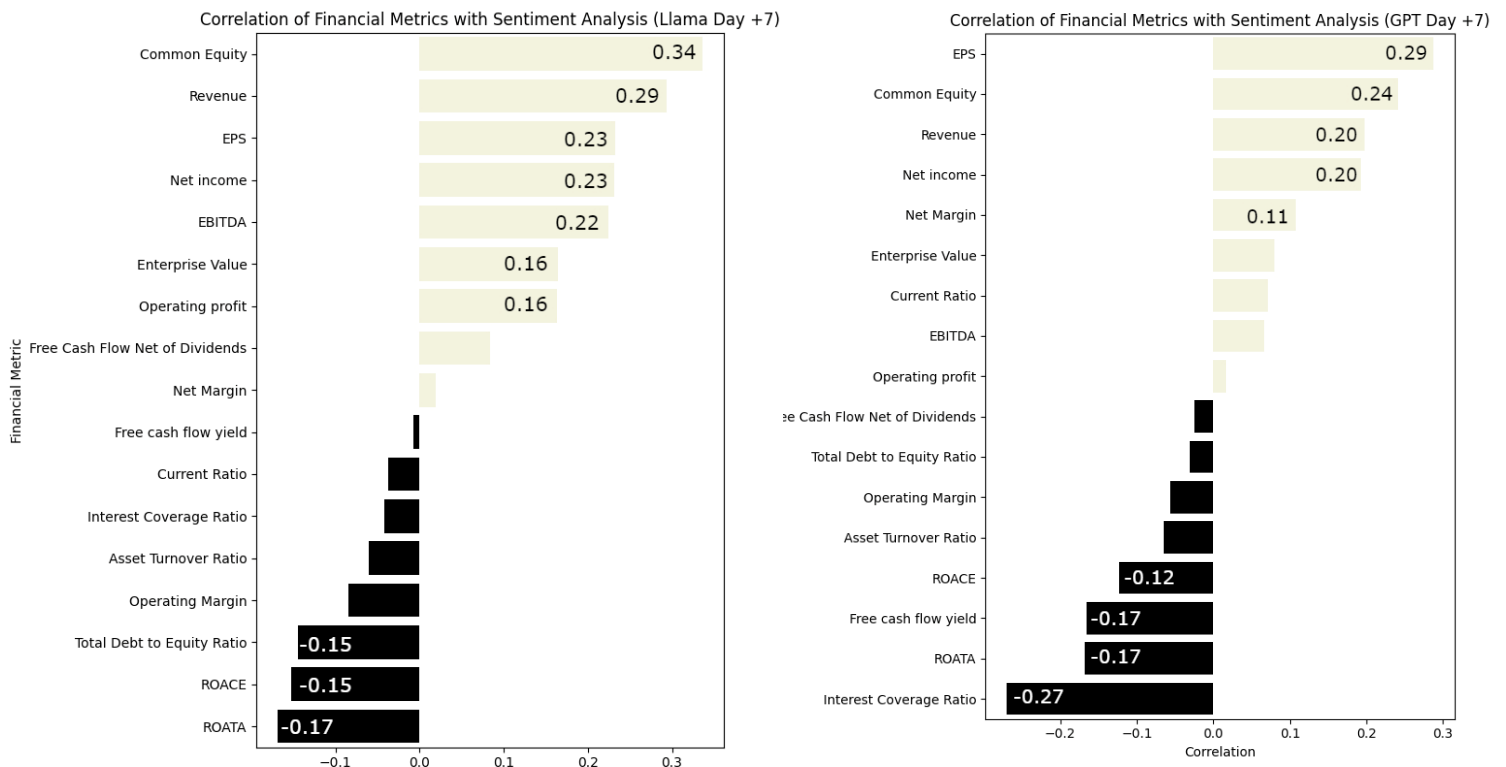


Figure 14: Significant correlations financial metrics and sentiment (Day 7) Llama 8B Instruct (left) sentiment and GPT 3.5 sentiment (right).

On Day +7, I analyzed correlations between financial metrics and multi-agent sentiment for both models. Both Llama 8B Instruct and GPT-3.5 showed notable positive relationships with EPS, Common Equity, Revenue, and Net Income. These metrics displayed weak positive correlations with multi-agent sentiment, suggesting a symmetrical relationship. The strengths of these correlations varied for per model. Llama 8B Instruct showed weak positive correlations with Enterprise Value and Operating Profit. These metrics had negligible correlations in GPT-3.5. Conversely, GPT-3.5 showed a weak positive correlation with Net Margin, which was negligible in Llama 8B Instruct. This highlights the differences in sentiment classification between the two models. Both models showed weak negative correlations with ROATA. There were notable differences in their negative correlations. Llama 8B Instruct showed weak negative correlations with Total Debt to Equity Ratio and ROACE. GPT-3.5 showed weak negative correlations with Interest Coverage Ratio, Free Cash Flow Yield, and ROACE. These negative relationships imply that increases in these

financial metrics are associated with a slight decline in market sentiment seven days after the release.

6 Discussion

In this section, each research question will be addressed based on the results presented in Chapter 4 and Chapter 5. The discussion will also cover the generalization and limitations of the research questions and the experiments conducted. By analyzing the findings, I aim to provide answers to the research questions, and acknowledge the constraints of the study. This section begins with answers to each research question, followed by a discussion of the study's limitations. The chapter concludes with potential directions for future research.

6.1 Addressing the Research Questions

This research examined the relationships between quarterly financial reports and short-term stock price trends. It explored how multi-agent sentiment analysis correlated with short-term stock price trends on three specific days: seven days before the financial release (Day -7), on the day of the release (Day 0), and seven days after the release (Day 7). Additionally, it investigated the relationship between quarterly financial metrics and multi-agent sentiment.

It is important to note that an observed correlation (i.e., association) does not assure that the relationship between 2 variables is causal (Schober et al., 2018). A significant correlation does not imply that changes in the variables caused changes in stock market trends. It simply states that the two variables move in the same direction. External factors could influence both, or the relationship could be coincidental. For the first research question of this study, I aimed to explore the relationship between financial metrics of quarterly financial reports and short-term stock trends in the 2022-2023 period.

RQ1: “How do specific quarterly financial metrics relate to stock market trends?”

Positive and negative correlations were observed for the financial metrics considered in this study. However, each significant correlation observed is interpreted as a weak correlation, for both the negative and the positive correlations. More specifically, Earnings per share, Interest Coverage Ratio, and Current Ratio indicated the same relationships with short-term stock market trends. This indicates that as these financial metrics increase, short-term stock market trends also tend to increase in the 2022-2023 period (positive relationship), and vice versa.

There is a negative relationship between Revenue, Asset Turnover Ratio, ROATA, EBITDA, Operating Profit, Enterprise Value, and short-term stock price trends in the 2022-2023 period. An increase in these variables is associated with a modest decrease in stock market trends. For specific correlation coefficients, see section 5.1. The second research question of this study aimed to uncover the relationships of multi-agent sentiment analysis and short-term stock trends in the 2022-2023 period.

RQ2: “How do multi-agent sentiment analysis of news headlines and tweets relate to stock market trends?”

The findings revealed that the relationship between sentiment analysis and stock price trends varied based on the timing of the sentiment capture and the LLM used.

- *Day -7*: Both LLMs show negligible negative correlations, indicating that sentiment a week before the financial release has no significant relationship with short-term stock price trends.
- *Day 0*: Llama 8B Instruct shows a weak positive correlation, suggesting that positive sentiment on the day of the financial release is associated with up trends. In contrast, GPT 3.5 shows a negligible correlation, indicating no relation with short-term stock trends on this day.
- *Day 7*: Both LLMs show negligible correlations, indicating that sentiment a week after the financial release does not significantly relate to short-term stock trends.

The analysis indicated that solely Llama 8B Instruct had a significant correlation on Day 0, which, however, is a weak positive correlation. The weak correlations of the qualitative and quantitative data with short-term stock trends could imply that other factors might play role. Considering quarterly financial metrics and qualitative data separately might not be sufficient, since these features provide only partial information (Lin et al., 2011). Additionally, stocks most sensitive to investor sentiment are those of younger, smaller, more volatile, unprofitable, non-dividend paying, distressed, or high-growth companies (Baker & Wurgler, 2007). This study considered nine companies in the S&P 500 index. No small-cap firms were included in

this sample. This may explain the weak and negligible correlations between multi-agent sentiment and short-term stock trends. The third research question in this study addresses the relationships of the financial metrics and the multi-agent sentiment analysis for both LLMs utilized.

RQ3: “What are the relationships between financial metrics and multi-agent sentiment around the time of quarterly financial report releases?”

The relationships between financial metrics and market sentiment around the time of quarterly financial report releases, as captured by multi-agent sentiment analysis, exhibit several key patterns based on correlation analysis. The specific correlation coefficients and financial metrics are discussed in section 5.4. These patterns indicate how market sentiment relates to various financial performance indicators before, during, and after the release of quarterly financial metrics:

Before the Release (Day -7): Both LLMs, Llama 8B Instruct and GPT 3.5, demonstrate weak positive and negative correlations with financial metrics such as Net Margin, EPS, and Net Income, and Interest Coverage Ratio. Weak positive correlations suggest that higher values of these financial metrics are weakly associated with positive sentiment, and vice versa. The weak correlation, however, indicates that other factors might be at play. For both LLMs utilized, there were weak negative correlations observed. Higher values of these financial metrics with weak correlations are associated with negative sentiment. These correlations suggest an inverse relationship between these financial metrics and sentiment. The symmetry of the correlations between the models, albeit with different strengths, suggest the alignment of sentiment classification of the models on Day -7.

On the Day of Release (Day 0): The correlations on the day of the financial release highlighted moderate positive correlations and weak negative correlations. With EPS showing the highest moderate positive correlation, and Net Income with a slightly weaker moderate positive correlation, for both models. The positive correlations indicate that there is a symmetrical relationship of the financial metrics with sentiment on the day of the release. The negative correlations of the financial metrics are the same as the negative correlations of Day -7, for

both models. This highlights the inverse relationship of these metrics with multi-agent sentiment.

After the Release (Day +7): One week post release, the analysis revealed weak positive correlations and weak negative correlations. Both models classified the three highest weak positive relationships for Common Equity, Revenue, and EPS. This suggests that as these financial metrics increased, sentiment became modestly more positive. However, for the weak negative correlations differences between the models existed. Both models showed a weak negative correlation with ROATA. But differences exist with Llama showing the weak negative correlations with Total Debt to equity ratio and ROACE. And GPT 3.5 showing weak negative correlations with Interest Coverage Ratio, and Free Cash Flow Yield. The negative correlations suggest weak inverse relationships of these financial metrics with market sentiment on Day 7.

Overall, the correlation analysis suggests that market sentiment, as captured by multi-agent sentiment analysis, is related to various financial metrics around the time of quarterly financial report releases. These relationships are evident before, during, and after the release of financial reports, highlighting a possible anticipatory and reactive nature of market sentiment in relation to financial performance indicators. The significant correlations on Day -7 could present potential considered anticipatory metrics by the market. Beaver., (1968) found that price and volume changes are most pronounced during the week of the earnings announcements, aligning with the moderate correlations presented on Day 0. The significant correlations on Day 0 indicate that earnings surprises in these metrics (Kormendi & Lipe., 1987) are relating to market sentiment. The persistence of earnings innovations impacts stock returns beyond the initial announcement, as investors adjust their expectations again for future earnings (Kormendi & Lipe, 1987). Connecting with the significant correlations on Day 7, it could that the market is continuing to process and react to these financial metrics even after the initial announcement. However, it is important to stay cautious about these findings since correlation does not imply causation.

6.2 Limitations

This research has several limitations that highlight aspects of improvement. All the data utilized is collected in a limited time period, namely in the quarters of 2022-2023. Due to

resource and time constraints, news headlines and tweets data are collected solely from Refinitiv Workspace. This limits the qualitative findings of this research to the data of the 41 news outlets utilized by Refinitiv and selected Twitter accounts by Refinitiv. Due to this, the sentiment analysis is based on a specific set of news headlines and tweets, which may not represent the entire spectrum of market sentiment. Additionally, sentiment can change rapidly, and the sentiment captured at specific times (Day -7, Day 0, Day 7) may not fully represent the ongoing sentiment dynamics. An important limitation is that Refinitiv Workspace did not include tweets in the news data for 2022; the platform only provided tweets from 2023. The findings may be specific to the market conditions during the 2022-2023 period and may not apply in different economic contexts or market environments. The study also considered a specific set of financial metrics. Other important factors that might influence market sentiment, such as geopolitical events or macroeconomic indicators, are not included in this analysis. Emphasizing the experiment, the multi-agent sentiment only encompasses Llama 3 8B Instruct and GPT 3.5 as key components. As the results have indicated, differences exist in each LLM utilized for multi-agent sentiment analysis, this limits the generalization of the results for LLMs. It is important to state the sentiment analysis conducted by the multi-agent system is not guaranteed to be without mistakes, as the accuracy of the sentiment analysis is not tested due to time constraints. It's also important to note that the relationships in this study are based on correlation analysis, meaning that while there are associations between financial metrics multi-agent sentiment, and short-term stock trends, this does not imply direct causation.

6.3 Recommendations for Future Research

This work can be extended into various venues of future research, that can test several other theories for enhancing the findings of this research. Future research can focus on several pathways, namely:

- *Compare Multi-Agent Sentiment Analysis with Traditional and Innovative Sentiment Analysis Tools:* Future research can explore if the sentiment analysis team employed in this research performs better than traditional or innovative sentiment analysis tools. This could involve a comparative study evaluating the accuracy and predictive power of different sentiment analysis methodologies.

- *Conduct Causal Analysis of the interplay of the variables utilized in this study:* Causal inference would allow researchers to determine whether changes in financial metrics actually causes changes in sentiment or vice versa. This research has opened 3 venues for causal analysis. This could involve techniques like Granger causality tests in time series data, instrumental variable analysis, or randomized control trials where feasible.
- *Compare the Multi-Agent Sentiment Analysis conversation patterns of this Research with Other Multi-Agent Sentiment Analysis conversation patterns:* This research could investigate the effectiveness of different multi-agent sentiment analysis frameworks. Comparing various frameworks may reveal strengths and weaknesses, leading to improved sentiment analysis techniques utilizing multi-agent conversations.
- *Predict Short-Term Market Trends with the Variables Employed in this Study:* Building on the correlation analysis, future research could develop predictive models using the identified variables to forecast short-term market trends. This could involve machine learning techniques to enhance the accuracy of predictions.
- *Longitudinal Study to Capture Longer-Term Trends:* Future studies could extend the temporal scope of the analysis to capture longer-term trends and the impact of different market conditions over time. This would provide a more comprehensive understanding of the dynamics between financial performance, market sentiment, and stock price movements.
- *Incorporate Additional Financial Metrics and Alternative Data Sources:* Expanding the set of financial metrics and incorporating alternative data sources such as social media sentiment, economic indicators, and geopolitical events could offer a more holistic view of the factors influencing stock market trends.

7 Conclusion

This study aimed to explore the relationships between quarterly financial earnings metrics, multi-agent sentiment analysis, and short-term stock trends. By employing a mixed-method approach, quantitative analysis of the variables, and qualitative sentiment analysis using a multi-agent conversation framework were conducted. The financial metrics analysed included Net Margin, EPS, Net Income, Revenue, and ratios such as Interest Coverage Ratio and Current Ratio. Sentiment data was collected and analysed on three specific days: seven days before the financial release (Day -7), on the day of the release (Day 0), and seven days after the release (Day 7). The multi-agent conversations utilized two LLMs, specifically Llama 8B Instruct and GPT-3.5, and was conducted on a corpus of 19,715 news headlines and tweets related to selected companies. Correlation analysis identified significant relationships between the variables, suggesting that sentiment analysis and certain financial metrics relate to market trends. However, considered separately, the relationships were weak, indicating the influence of multiple factors on short-term stock market trends. Consecutively, significant correlations were found between certain financial metrics and sentiment. The strongest correlations were observed particularly on the release day of quarterly financial reports, highlighting the unique relationships of the financial metrics and sentiment around the release dates. The study also found that the choice of LLM affects the sentiment scores assigned by the sentiment analysis team on specific days, underscoring the difference of sentiment analysis conducted by each distinct model. The academic value of this research contributes to a deeper understanding of the interplay of the variables employed in this study, creating new possible research venues. Possible future research venues encompass stock trend prediction studies, causal studies, and performance comparison studies. Overall, the findings underscore the complexity of market dynamics. As generative AI technologies become more accessible, their integration into financial analysis will likely accelerate, transforming the industry into an AI-driven domain.

References

- Abdullah, T., & Ahmet, A. (2022). Deep learning in sentiment analysis: Recent architectures. *ACM Computing Surveys*, 55(8), 1–37. <https://doi.org/10.1145/3548772>
- Baker, M., & Wurgler, J. (2007). *Investor Sentiment in the Stock Market*.
<https://doi.org/10.3386/w13189>
- Barberis, N., & Thaler, R. (2002). *A Survey of Behavioral Finance*.
<https://doi.org/10.3386/w9222>
- Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of Accounting Research*, 6, 67. <https://doi.org/10.2307/2490070>
- Bryman, A. (2003). *Quantity and Quality in Social Research*.
<https://doi.org/10.4324/9780203410028>
- Brynjolfsson, E., Li, D., & Raymond, L. (2023). *Generative AI at Work*.
<https://doi.org/10.3386/w31161>
- Bukovina, J. (2016). Social media big data and Capital Markets—an overview. *Journal of Behavioral and Experimental Finance*, 11, 18–26.
<https://doi.org/10.1016/j.jbef.2016.06.002>
- Chambers, A. E., & Penman, S. H. (1984). Timeliness of reporting and the stock price reaction to earnings announcements. *Journal of Accounting Research*, 22(1), 21.
<https://doi.org/10.2307/2490700>
- Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3), 1–45. <https://doi.org/10.1145/3641289>
- Chang, P.-C., Wang, D., & Zhou, C. (2012). A novel model by evolving partially connected neural network for stock price trend forecasting. *Expert Systems with Applications*, 39(1), 611–620. <https://doi.org/10.1016/j.eswa.2011.07.051>

- Chowdhary, K. R. (2020). *Fundamentals of Artificial Intelligence*. Springer (India) Private Limited.
- Cutler, D., Poterba, J., & Summers, L. (1988). *What Moves Stock Prices?*
<https://doi.org/10.3386/w2538>
- Dechow, P. M., Sloan, R. G., & Zha, J. (2013). Stock prices & earnings: A history of research. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2347193>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805v2*.
<https://doi.org/> <https://doi.org/10.48550/arXiv.1810.04805>
- Du, K., Xing, F., Mao, R., & Cambria, E. (2024). Financial sentiment analysis: Techniques and applications. *ACM Computing Surveys*, 56(9), 1–42.
<https://doi.org/10.1145/3649451>
- Du, Y., Li, S., Torralba, A., Tenenbaum, J. B., & Mordatch, I. (2023). Improving Factuality and Reasoning in Language Models through Multiagent Debate. *arXiv:2305.14325v1*.
- Fama, E. F. (1970). Efficient Capital Markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383. <https://doi.org/10.2307/2325486>
- Giray, L. (2023). Prompt engineering with CHATGPT: A guide for academic writers. *Annals of Biomedical Engineering*, 51(12), 2629–2633. <https://doi.org/10.1007/s10439-023-03272-4>
- Google. (n.d.). Google Gemini. Retrieved June 6, 2024, from <https://gemini.google.com/>
- Groqcloud. Groq Cloud. (n.d.). Retrieved May 05, 2024, from <https://console.groq.com/docs/quickstart>
- Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., & Zhang, X. (2024). Large Language Model based Multi-Agents: A Survey of Progress and Challenges. *arXiv:2402.01680v1*. <https://doi.org/>
<https://doi.org/10.48550/arXiv.2402.01680>

- Hugging Face. (n.d.). Meta Llama 3-8B Instruct. Hugging Face. Retrieved May 24, 2024, from <https://huggingface.co/meta-llama/Meta-Llama-3-8B-Instruct>
- Jaiswal, S., Srivastava, S., Garg, S., & Singh, P. (2023). Effect of news headlines on Gold Price prediction using NLP and deep learning. *2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1)*. <https://doi.org/10.1109/icaia57370.2023.10169488>
- Kasneji, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneji, G. (2023). Chatgpt for good? on opportunities and challenges of large language models for Education. *Learning and Individual Differences, 103*, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kearney, C., & Liu, S. (2013). Textual sentiment analysis in finance: A survey of methods and models. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2213801>
- Kormendi, R., & Lipe, R. (1987). Earnings Innovations, earnings persistence, and stock returns. *The Journal of Business, 60*(3), 323. <https://doi.org/10.1086/296400>
- Koshiyama, A., Firoozye, N., & Treleaven, P. (2020). Algorithms in future Capital Markets. *Proceedings of the First ACM International Conference on AI in Finance*. <https://doi.org/10.1145/3383455.3422539>
- Li, Y., Wang, S., Ding, H., & Chen, H. (2023). Large language models in Finance: A Survey. *4th ACM International Conference on AI in Finance*. <https://doi.org/10.1145/3604237.3626869>
- Lin, M.-C., Lee, A. J., Kao, R.-T., & Chen, K.-T. (2011). Stock price movement prediction using representative prototypes of financial reports. *ACM Transactions on Management Information Systems, 2*(3), 1–18. <https://doi.org/10.1145/2019618.2019625>
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Springer.

- London Stock Exchange Group. (n.d.). Workspace. London Stock Exchange Group. Retrieved March 23, 2024 from <https://www.lseg.com/en/data-analytics/products/workspace#working-with-workspace>
- Lu, Y., Bartolo, M., Moore, A., Riedel, S., & Stenetorp, P. (2022). Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. <https://doi.org/10.18653/v1/2022.acl-long.556>
- Magnusson, C., Arppe, A., Eklund, T., Back, B., Vanharanta, H., & Visa, A. (2005). The language of quarterly reports as an indicator of change in the company's financial status. *Information & Management*, 42(4), 561–574. <https://doi.org/10.1016/j.im.2004.02.008>
- Markowitz, H. M. (1991). Foundations of portfolio theory. *The Journal of Finance*, 46(2), 469. <https://doi.org/10.2307/2328831>
- McKinney, W. (2010). Data Structures for Statistical Computing in python. *Proceedings of the Python in Science Conference*. <https://doi.org/10.25080/majora-92bf1922-00a>
- Microsoft. (2024, March 3). AutoGen Documentation. Microsoft. Retrieved May 01, 2024 from <https://microsoft.github.io/autogen/>
- Microsoft Corporation. (n.d.). Visual Studio Code. Retrieved May 05, 2024, from <https://code.visualstudio.com/>
- OpenAI. (n.d.-a). ChatGPT. Retrieved June 6, 2024, from <https://chatgpt.com/?oai-dm=1>
- OpenAI. (n.d.-b). *OpenAI API documentation*. OpenAI. Retrieved May 05, 2024, from <https://platform.openai.com/docs/>
- Python Software Foundation. (2023). *Python 3.10.0 documentation*. Python.org. Retrieved May 15, 2024, from <https://docs.python.org/3.10/reference/index.html>
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5), 1763–1768. <https://doi.org/10.1213/ane.0000000000002864>

- Shleifer, A., & Summers, L. H. (1990). The Noise Trader Approach to finance. *Journal of Economic Perspectives*, 4(2), 19–33. <https://doi.org/10.1257/jep.4.2.19>
- Sidogi, T., Mbuva, R., & Marwala, T. (2021). Stock price prediction using sentiment analysis. *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. <https://doi.org/10.1109/smc52423.2021.9659283>
- Singh, U., Abhishek, K., & Azad, H. K. (2024). A survey of cutting-edge multimodal sentiment analysis. *ACM Computing Surveys*, 56(9), 1–38. <https://doi.org/10.1145/3652149>
- Song, C.-H., Xiao, X., Zhang, B., & Xia, S.-T. (2023). Follow the will of the market: A context-informed drift-aware method for stock prediction. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. <https://doi.org/10.1145/3583780.3614886>
- Sun, X., Wang, G., Zhang, T., Li, J., Wu, F., Wang, S., Zhang, S., & Li, X. (2023). Sentiment Analysis through LLM Negotiations. *arXiv:2311.01876v1 [Cs.CL]* . <https://doi.org/https://doi.org/10.48550/arXiv.2311.01876>
- Thompson, A. (2024, March 25). *Digital 2024: 5 billion Social Media Users*. We Are Social UK. <https://wearesocial.com/uk/blog/2024/01/digital-2024-5-billion-social-media-users/>
- Vaswani, A., Shazeer, N., Parmar, N., Polosukhin, I., Kaiser, Ł., Gomez, A. N., Jones, L., & Uszkoreit, J. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. <https://doi.org/10.48550/arXiv.1706.03762>
- 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., & Wen, J. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6). <https://doi.org/10.1007/s11704-024-40231-1>
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering

with ChatGPT. *arXiv:2302.11382 [Cs.SE]*. <https://doi.org/https://doi.org/10.48550/arXiv.2302.11382>

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Wang, C., Burger, D., White, R. W., Awadallah, A., Liu, J., Zhang, S., Zhang, X., Jiang, L., Zhu, E., & Li, B. (2023). AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. *arXiv:2308.08155v2*. <https://doi.org/https://doi.org/10.48550/arXiv.2308.08155>

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Cao, Y., & Narasimhan, K. (2023). React: Synergizing Reasoning and Acting in Language Models. *arXiv:2210.03629v3 [Cs.CL]*. <https://doi.org/https://doi.org/10.48550/arXiv.2210.03629>

Zhang, W., Deng, Y., Bing, L., Pan, S. J., & Liu, B. (2023). Sentiment Analysis in the Era of Large Language Models: A Reality Check. *arXiv:2305.15005*. <https://doi.org/https://doi.org/10.48550/arXiv.2305.15005>

Appendix A: Data

A.1 Data specifications table

The data specifications table, presented in the table below, serves as a guide to the foundational elements of my data collection and analysis process. It outlines the subject area under computer science, with a specific focus on financial data analysis leveraging the Refinitiv Workspace, enhanced through Python scripting and Visual Studio Code (VSCoDe). This table details the types of data collected, the methodologies employed for data acquisition, the formats in which data were processed, and the parameters considered during collection. It also provides information on how to access these datasets and their potential for reuse, underscoring the structured and methodical approach taken in gathering and analyzing financial data for insightful market analysis.

Subject	Computer Science, Finance
Specific subject area	Financial Data analysis using Refinitiv workspace.
Type of data	Table
How data were acquired	Data were acquired through Refinitiv workspace, filtered and homogenized with Python3 in the SDK Visual Studio Code (VSCoDe)
Data format	Raw Filtered Analyzed
Parameters for data collection	Parameters included are financial performance metrics per quarter, news headlines, and social media posts. For news data, <company name> and English were keywords utilized to source news and tweets related to the companies considered. The data are acquired through Refinitiv Workspace, accessible through a subscription. All data timeframes range from the 31-03-2022 up until 31-12-2023. Python Scripts were tailored to extract specific data points relevant to these parameters.

Description of data collection	Data are obtained through Refinitiv Workspace. The data are exported to excel files through the user interface the app provides. The news headlines and tweets are copied manually into excel files, since Refinitiv Workspace would now allow for direct transportation of the data into excel files. Financial summaries and price data were directly transportable.
Data source location	https://www.lseg.com/en/data-analytics/products/workspace https://developers.lseg.com/en/api-catalog/eikon/eikon-data-api/quick-start
Data accessibility	Data access is granted through acquiring a license to Refinitiv workspace. Otherwise, accessibility is restricted. Access is granted by emailing lis-financialdatasupport@tilburguniversity.edu .

A.2 Data Value

- This dataset integrates qualitative sentiment from news and social media on three specific days (7 days before, on the day of, and 7 days after the release of quarterly financial summaries). Additionally, quantitative metrics of quarterly financial reports are incorporated. This combination provides a perspective on how these factors influence stock market trends. The stock market trends are calculated using moving averages of price data, which span from 7 days before the release of the quarterly financial summaries to 7 days after the release.
- Financial analysts, AI researchers, and scholars in economics and computational linguistics can utilize this dataset to develop sophisticated predictive models, enhance financial decision-making processes, and refine sentiment analysis methodologies. Additionally, professionals in the financial sector can leverage this data to consider the most impactful metrics relating to a company. Incorporating qualitative data and quantitative data to make informed decisions.
- The dataset serves for future research avenues, including the synergy between sentiment analysis and financial metrics for refined AI predictions, understanding

investor psychology and company earnings (causal analysis), and applying insights to broader markets.

A.3 Metrics used in Q-file

Metric	Description	Relevance for the Study
Revenue	Total income generated from business activities.	Essential for assessing the company's sales performance.
Operating Profit	Profit from business operations before accounting for non-operating expenses such as interest and taxes.	Highlights the operational efficiency of the company.
EBITDA	Earnings before interest, taxes, depreciation, and amortization.	Highlights the operational efficiency of the company.
Net Income	Total earnings remaining after all expenses and taxes have been deducted.	Reflects the overall profitability of the company.
Common Equity	Equity owned by common shareholders.	Used to assess the company's financial stability and the valuation of its equity
Free Cash Flow Net of Dividends	Cash a company generates after cash outflows to support operations, deducted by dividends paid to shareholders.	Shows how much cash remains after all capital expenditures and dividend payments are covered.
EPS (Earnings per share)	Profit attributed to each individual share of stock.	Directly relates to the profitability of the company per share,
Operating Margin	Ratio of operating income to net sales.	Indicates how efficiently a company is converting sales into profits
Net Margin	Net income as a percentage of revenue.	Measures profitability efficiency by showing how much of the revenue is converted into actual profit.
Free Cash Flow Yield	Free cash flow relative to its share price.	Helps investors assess the attractiveness of a stock by showing how much cash can be extracted from each share relative to its price.

ROACE	Return on average capital employed.	Evaluates the profitability and efficiency in using capital, influencing investment decisions.
ROATA	Return on average total assets.	Shows how effective management is at using its assets to generate earnings.
Total Debt to Equity Ratio	Ratio of total debt to shareholders' equity.	Indicates the company's leverage and financial health, which can affect stock stability and investor trust.
Interest Coverage Ratio	Ability of a company to pay interest on its outstanding debt.	A measure of a firm's financial strength and ability to meet its debt obligations, influencing credit ratings and stock prices.
Enterprise Value	Market capitalization plus debt, minority interest, and preferred shares, minus total cash and cash equivalents.	Provides a comprehensive valuation that helps compare companies with different capital structures.
Asset Turnover Ratio	The ratio of revenue over assets.	Measures the company's ability to use its assets to generate sales, important for assessing operational efficiency.
Current Ratio	A liquidity ratio that measures a company's ability to pay short-term obligations.	Useful for evaluating financial health and operational efficiency, impacting investor sentiment and market reactions to financial reports.

A.4 Quantitative Data File Snapshot

Company	Quarter	Revenue	Operating profit	EBITDA	Net income	Common Equity	Free Cash Flow	EPS	Operating Margin	Net Margin	Free cash flow yield	ROACE	ROATA	Total Debt to Equity Ratio	Interest Coverage Ratio	Enterprise Value	Asset Turnover Ratio	Current Ratio	
GOOGL	31/03/2022	68.011	20.094	23.876	16.436	255.419	12.594	1,23	29,6%	24,2%	3,6%	30,8%	21,8%		5,9%	242,1	1.723.069	0,79	2,87
AMZN	31/03/2022	116.444	3.918,0	12.896	-3.844,0	134.001	-17.741	-0,38	3,4%	-3,3%	-1,5%	18,1%	5,8%		58,3%	8,30	1.669.612	1,30	0,96
APPL	31/03/2022	97.278	29.979	32.716	25.010	67.399	22.057	1,52	30,8%	25,7%	3,7%	149,3%	29,6%		178,0%	43,38	2.919.803	1,12	0,93
BRKa	31/03/2022	70.843	9.630,0	12.333	5.705,0	508.141	3.739,0	3.783,8	13,6%	8,1%	2,9%	17,5%	9,2%		23,2%	9,31	801.669	0,30	2,35
LLY	31/03/2022	7.810,0	2.569,9	3.005,6	1.902,9	9.330,8	756,5	2,10	32,9%	24,4%	2,2%	75,5%	13,1%		174,5%	30,27	286.795	0,63	1,27
META	31/03/2022	27.908	8.524,0	10.684	7.465,0	123.228	8.635,0	2,72	30,5%	26,8%	6,3%	29,1%	22,8%		0,5%	2.131,0	561.966	0,73	2,81
MSFT	31/03/2022	49.360	20.364	24.137	16.728	162.924	15.401	2,22	41,3%	33,9%	2,4%	48,7%	22,2%		39,6%	40,49	2.256.592	0,59	1,99
NVDA	31/03/2022	7.643,0	2.970,0	3.279,0	3.003,0	26.612	2.659,0	1,18	38,9%	39,3%	1,4%	44,8%	26,7%		41,1%	48,69	560.738	0,74	6,65
TSLA	31/03/2022	18.756	3.603,0	4.483,0	3.280,0	34.085	2.214,0	0,95	19,2%	17,5%	0,6%	29,4%	14,2%		13,8%	72,06	1.104.488	1,05	1,35
GOOGL	30/06/2022	69.685	19.453	23.352	16.002	255.419	12.594	1,21	27,9%	23,0%	4,4%	29,2%	20,9%		5,8%	234,4	1.327.428	0,81	2,81
AMZN	30/06/2022	121.234	3.407,0	13.001	-2.028,0	131.402	-6.759,0	-0,20	2,8%	-1,7%	-2,7%	9,4%	3,0%		67,4%	5,83	1.109.345	1,25	0,95
APPL	30/06/2022	82.959	23.076	25.881	19.442	58.107	16.979	1,20	27,8%	23,4%	4,6%	162,8%	29,9%		206,0%	32,09	2.364.253	1,16	0,86
BRKa	30/06/2022	76.201	12.494	15.204	-43.242	461.229	4.789,0	-29.663	16,4%	-56,8%	3,4%	2,6%	1,4%		25,4%	11,49	624.105	0,32	2,32
LLY	30/06/2022	6.488,0	1.650,5	1.999,4	952,5	8.544,7	-514,0	1,05	25,4%	14,7%	1,7%	75,9%	12,0%		194,2%	20,33	322.262	0,61	1,10
META	30/06/2022	28.822	8.358,0	10.341	6.687,0	125.767	4.624,0	2,46	29,0%	23,2%	8,0%	25,5%	19,8%		0,5%	2.089,5	396.518	0,70	2,52
MSFT	30/06/2022	51.865	20.534	24.513	16.740	166.542	13.126	2,23	39,6%	32,3%	3,4%	47,2%	20,8%		38,8%	41,40	1.876.905	0,57	1,78
NVDA	30/06/2022	8.288,0	3.377,0	3.711,0	1.618,0	26.320	1.270,0	0,64	40,8%	39,5%	1,7%	42,0%	24,9%		49,6%	455.029	0,78	5,32	
TSLA	30/06/2022	16.934	2.606,0	3.358,0	2.269,0	36.376	621,0	0,65	15,4%	13,4%	0,9%	31,1%	15,5%		11,9%	59,23	684.723	1,09	1,43
GOOGL	30/09/2022	69.092	17.135	21.181	13.910	253.626	16.077	1,06	24,8%	20,1%	4,8%	26,9%	19,0%		5,9%	169,7	1.142.539	0,80	2,52
AMZN	30/09/2022	127.101	2.690,0	12.894	2.872,0	137.489	-4.974,0	0,28	2,1%	2,3%	-2,3%	8,8%	2,8%		6,3%	4,36	1.180.463	1,24	0,94
APPL	30/09/2022	90.146	24.894	27.759	20.721	50.672	17.135	1,29	27,6%	23,0%	4,5%	175,5%	28,4%		237,0%	30,10	2.489.288	1,12	0,88
BRKa	30/09/2022	76.904	9.884,0	12.612	-2.698,0	455.430	7.604,0	-1.907,4	12,9%	-3,5%	3,3%	-0,2%	0,0%		25,1%	9,05	607.133	0,32	2,35
LLY	30/09/2022	6.941,6	1.945,4	2.308,3	1.451,7	10.070	682,4	1,61	28,0%	20,9%	1,8%	67,7%	12,6%		156,3%	23,87	320.481	0,61	1,13
META	30/09/2022	27.714	5.664,0	7.839,0	4.395,0	124.094	317,0	1,64	20,4%	15,9%	7,1%	22,4%	16,6%		8,5%	1.416,0	333.374	0,68	2,57
MSFT	30/09/2022	50.122	21.518	24.308	17.556	173.566	12.294	2,35	42,9%	35,0%	3,6%	42,9%	20,1%		36,4%	43,04	1.692.725	0,58	1,84
NVDA	30/09/2022	6.704,0	674,0	1.052,0	656,0	23.851	737,0	0,26	10,1%	9,8%	1,4%	34,4%	18,9%		45,9%	10,37	447.987	0,72	3,62
TSLA	30/09/2022	21.454	3.688,0	4.644,0	3.331,0	39.851	3.297,0	0,95	17,2%	15,5%	1,0%	33,4%	17,0%		8,7%	92,20	814.872	1,13	1,46
GOOGL	31/12/2022	76.048	18.160	22.362	13.624	256.144	16.019	1,05	23,9%	17,9%	5,2%	23,6%	16,6%		5,9%	201,8	1.046.241	0,78	2,38
AMZN	31/12/2022	149.204	3.333,0	16.478	278,0	146.043	12.581	0,03	2,2%	0,2%	-2,0%	-1,9%	-0,6%		63,6%	4,80	879.811	1,16	0,94
APPL	31/12/2022	117.154	36.016	38.932	29.998	56.727	26.450	1,88	30,7%	25,6%	4,6%	147,9%	26,2%		195,9%	35,91	2.126.697	1,06	0,94
BRKa	31/12/2022	78.072	7.880,0	10.638	18.237	473.424	5.754,0	12.355	10,1%	23,4%	3,2%	-4,7%	-2,3%		25,5%	6,92	684.337	0,32	2,51
LLY	31/12/2022	7.301,8	2.114,6	2.489,6	1.937,7	10.650	139,7	2,15	29,0%	26,5%	1,4%	63,6%	12,7%		150,7%	25,17	361.765	0,58	1,05
META	31/12/2022	32.165	11.010	13.387	6.163,0	125.713	5.468,0	2,03	34,2%	19,2%	5,7%	19,7%	14,1%		8,4%	2.752,5	285.427	0,66	2,20
MSFT	31/12/2022	52.747	21.570	25.218	16.425	183.136	-167,0	2,20	40,9%	31,1%	3,3%	39,3%	19,1%		34,8%	44,02	1.751.924	0,58	1,93
NVDA	31/12/2022	5.931,0	791,0	1.197,0	680,0	21.349	-238,0	0,27	13,3%	11,5%	1,4%	26,4%	14,7%		51,3%	12,17	342.274	0,70	3,39
TSLA	31/12/2022	24.318	3.935,0	4.890,0	3.707,0	44.704	1.420,0	1,07	16,2%	15,2%	1,8%	33,6%	17,4%		6,8%	119,2	371.080	1,13	1,53
GOOGL	31/03/2023	69.787	19.973	23.159	15.051	260.894	17.220	1,18	28,6%	21,6%	4,6%	22,8%	16,1%		5,8%	249,7	1.230.047	0,78	2,35
AMZN	31/03/2023	127.358	4.774,0	15.897	3.172,0	154.526	-9.419,0	0,31	3,8%	2,5%	-0,8%	3,0%	1,0%		58,5%	5,80	1.085.543	1,20	0,92
APPL	31/03/2023	94.836	28.318	31.216	24.160	62.158	21.994	1,52	29,9%	25,5%	3,7%	145,6%	27,6%		176,4%	30,45	2.646.533	1,13	0,94
BRKa	31/03/2023	85.393	10.524	13.575	35.757	504.554	4.980,0	24.377	12,3%	41,9%	3,4%	1,4%	0,8%		24,1%	8,64	680.371	0,32	2,44
LLY	31/03/2023	6.960,0	1.599,0	1.961,3	1.344,9	11.190	-190,1	1,49	23,0%	19,3%	1,2%	55,4%	11,4%		167,2%	15,55	342.374	0,55	1,30
META	31/03/2023	28.645	7.848,0	10.375	5.709,0	124.795	1.156,0	2,20	27,4%	19,9%	3,0%	18,5%	13,2%		8,5%	142,7	517.060	0,67	2,07
MSFT	31/03/2023	52.857	22.352	25.901	18.299	194.683	12.775	2,45	42,3%	34,6%	2,7%	38,6%	19,1%		33,2%	45,06	2.104.333	0,57	1,91
NVDA	31/03/2023	6.051,0	1.485,0	1.911,0	1.414,0	22.101	1.641,0	0,57	24,5%	23,4%	0,8%	17,9%	10,2%		49,6%	23,20	498.636	0,63	3,52
TSLA	31/03/2023	23.329	2.664,0	3.710,0	2.539,0	48.054	440,0	0,73	11,4%	10,9%	0,8%	28,7%	15,5%		5,5%	91,86	638.961	1,13	1,57

A.5 Qualitative Data File Snapshot

	A	B	C	D	E
1		date	news	llama	gpt
2	0	02/04/20	Users with iPhone, iPad and Mac must urgently download latest security patch	-0,5	0
3	1	02/04/20	Intellectual Property India Publishes Patent Application for 'User Interface Camera Effects' Filed by Apple Inc.	0	0
4	2	02/04/20	Intellectual Property India Publishes Patent Application for 'User Interfaces For Viewing Live Video Feeds And Recorded Video' Filed by Apple Inc.	0	0
5	3	02/04/20	Intellectual Property India Publishes Patent Application for 'Devices, Methods, And Graphical User Interfaces For Navigating Between User Interface...	0	0
6	4	02/04/20	Intellectual Property India Publishes Patent Application for 'Reference Circuit For Metrology System' Filed by Apple Inc.	0	0
7	5	02/04/20	Q1 2022 Report: Muhlenkamp & Co reduces exposure to Apple by 30.2%	-1	-1
8	6	02/04/20	Q1 2022 Report: Park National trims its losses in Apple	-0,5	-0,5
9	7	02/04/20	Q1 2022 Report: New Mexico Educational Retirement Board trims its losses in Apple	-0,5	-0,5
10	8	02/04/20	iPad Air 5 Features Pull Tabs for Easier Battery Replacements	0	1
11	9	02/04/20	First Android smartphone created with Apple-exclusive technology	0	1
12	10	02/04/20	In Apple v Epic, 9th Circuit Should Remember that Antitrust Forbids Enhancing, not Exercising, Market Power	-1	0
13	11	02/04/20	Q1 2022 Report: Muhlenkamp & Co reduces exposure to Apple by 30.2% Saturday April 02, 2022	-0,5	-1
14	12	02/04/20	Q1 2022 Report: Park National trims its losses in Apple Saturday April 02, 2022	-0,5	-0,5
15	13	02/04/20	Q1 2022 Report: New Mexico Educational Retirement Board trims its losses in Apple Saturday April 02, 2022	0	-0,5
16	14	02/04/20	Intellectual Property India Publishes Patent Application for 'Method For Multiplexing Between Power Supply Signals For Voltage Limited Circuits' Fi...	-0,5	0
17	15	02/04/20	Why Some Apple Workers Are Threatening To Quit Over A Policy That Takes Effect On April 11	0	0
18	16	02/04/20	Intellectual Property India Publishes Patent Application for 'Sensor Assemblies For Electronic Devices' Filed by Apple Inc.	0	0,5
19	17	02/04/20	Report: Apple Is Declaring Several MacBooks Obsolete	0	-1
20	18	02/04/20	Apple @ Work: Using web apps on macOS is streamlined with site specific browser applications	0	1
21	19	02/04/20	Apple Inc Files Patent Application for on the Frame Structure Design for Single Carrier Waveform	0	0,5
22	20	02/04/20	Apple Inc Files Patent Application for Method and Apparatus for New Radio Unlicensed (NR-U) Radio Link Monitoring	0	0,5
23	21	02/04/20	Apple Inc Files Patent Application for Reception (Rx) Beam Sweep Scale Factor	0	0,5
24	22	02/04/20	Apple Inc Files Patent Application for Grant Based and Configured Grant Based PUSCH Transmissions in an Unlicensed Spectrum	0	0,5
25	23	02/04/20	Top Stories: iOS 15.4.1 Released, Studio Display Teardown, and More	0	1
26	24	02/04/20	Apple M1 Max vs. M1 Ultra, Tested: Apple's Top Desktop Chips Face Off	-1	0,5
27	25	02/04/20	Kaizen Financial Strategies Trims Position in Apple Inc. (NASDAQ:AAPL)	0	0
28	26	02/04/20	New MacBook Air could debut later in 2022 to further spark Mac sales	1	0,5
29	27	02/04/20	Apple Inc. (NASDAQ:AAPL) is Narwhal Capital Management's Largest Position	1	1
30	28	02/04/20	Apple Inc. (NASDAQ:AAPL) Shares Bought by Lincoln National Corp	1	1
31	29	02/04/20	Austin Asset Management Co Inc Invests \$3.37 Million in Apple Inc. (NASDAQ:AAPL)	-1	1
32	30	02/04/20	Optas LLC Reduces Stock Holdings in Apple Inc. (NASDAQ:AAPL)	1	-1
33	31	02/04/20	Austin Asset Management Co Inc Buys Shares of 18,963 Apple Inc. (NASDAQ:AAPL)	-1	1
34	32	02/04/20	Apple Inc. (NASDAQ:AAPL) Stake Decreased by FLC Capital Advisors	1	1

A.6 Final Data File Snapshot

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
1	Company	Quarter	Revenue	Operating profit	EBITDA	Net income	Common Equity	Cash Flow Net of Div	EPS	Operating Margin	Net Margin	cash flow	ROACE	ROATA	Debt to Equity	Coverage	Enterprise Val.	Turnover	Current Ratio	Time-lane	Time-lane	Time-lane	Time-lane	Time-lane	Time-lane	Time-lane	Price Trend
2	GOOGL	31/03/2022	68011000000	20094000000	23876000000	16436000000	2,55419E+11	12594000000	1,23	0,2955	0,2417	0,0362	0,308	0,2179	0,0589	242,1	1,7231E+12	0,79	2,87	0,183333	-0,01111	-0,0098	0,183333	0,166667	0,147059	-1	
3	AMZN	31/03/2022	1,16444E+11	39180000000	12896000000	-3844000000	1,34001E+11	-17741000000	-0,38	0,0336	-0,033	-0,0147	0,1805	0,0584	0,5834	8,3	1,6696E+12	1,3	0,96	0,15873	0,007752	0,058219	0,190476	0,131783	0,171233	-1	
4	APPL	31/03/2022	97278000000	29979000000	32716000000	25010000000	67399000000	22057000000	1,52	0,3082	0,2571	0,0365	1,4927	0,2964	1,7802	43,38	2,9198E+12	1,12	0,93	0,019048	0,06689	-0,07018	0,174419	0,197324	0,219048	1	
5	BRKa	31/03/2022	70843000000	9630000000	12333000000	5705000000	5,08141E+11	3739000000	3783,81	0,1359	0,0805	0,0292	0,1753	0,0915	0,2315	9,31	8,0167E+11	0,3	2,35	0,115385	0,016667	0,347826	0,192308	0,233333	0,398551	-1	
6	LLY	31/03/2022	78100000000	25699000000	30056000000	19029000000	93308000000	7565000000	2,1	0,3291	0,2436	0,0224	0,7553	0,1307	1,7447	30,27	2,868E+11	0,63	1,27	0,160714	-0,29167	0,291667	0,017857	-0,02083	0,28125	1	
7	META	31/03/2022	27908000000	8524000000	10684000000	7465000000	1,23228E+11	8635000000	2,72	0,3054	0,2675	0,0634	0,2907	0,2279	0,0049	2131	5,6197E+11	0,73	2,81	-0,25333	-0,23853	-0,17157	-0,03333	-0,1789	-0,08824	1	
8	MSFT	31/03/2022	49360000000	20364000000	24137000000	16728000000	1,62924E+11	15401000000	2,22	0,4126	0,3389	0,0238	0,4872	0,2218	0,3957	40,49	2,2566E+12	0,59	1,99	0,293651	0,104167	0,253012	0,187831	0,219907	0,292169	-1	
9	NVDA	31/03/2022	76430000000	29700000000	32790000000	30030000000	266120000000	26590000000	1,18	0,3886	0,3929	0,014	0,4483	0,2673	0,4113	48,69	5,6074E+11	0,74	6,65	0,003546	-0,01852	-0,02941	0,4	-0,16667	0,428571	1	
10	TSLA	31/03/2022	187560000000	36030000000	44830000000	32800000000	340850000000	22140000000	0,95	0,1921	0,1749	0,0057	0,2942	0,1422	0,1377	72,06	1,1045E+12	1,05	1,35	0,535088	0,089286	0,1	0,333333	0,071429	0,194444	1	
11	GOOGL	30/06/2022	69685000000	19453000000	23352000000	16002000000	2,55419E+11	12594000000	1,21	0,2792	0,2296	0,0439	0,2922	0,2086	0,058	234,37	1,3274E+12	0,81	2,81	0,176471	-0,12903	0,511111	0,254902	-0,28226	0,333333	0	
12	AMZN	30/06/2022	1,21234E+11	34070000000	13001000000	-2028000000	1,31402E+11	-6759000000	-0,2	0,0281	-0,0167	-0,0273	0,0943	0,0298	0,6737	5,83	1,1093E+12	1,25	0,95	0,026432	-0,10938	0,210714	0,044053	-0,09766	0,246429	1	
13	APPL	30/06/2022	82959000000	23076000000	25881000000	19442000000	58107000000	16979000000	1,2	0,2782	0,2344	0,0462	1,6282	0,2991	2,0598	32,09	2,3643E+12	1,16	0,86	0,224057	0,011236	-0,07018	0,339623	0,337079	0,078947	0	
14	BRKa	30/06/2022	76201000000	12494000000	15204000000	-43242000000	4,61229E+11	4789000000	-29662,5	0,164	-0,5675	0,0343	0,026	0,0144	0,2536	11,49	6,2411E+11	0,32	2,32	-0,2	-0,22	0,078125	-0,23333	-0,36	0,21875	1	
15	LLY	30/06/2022	64880000000	16505000000	19994000000	9525000000	85447000000	-5140000000	1,05	0,2544	0,1468	0,0168	0,7594	0,12	1,9417	30,23	2,2226E+11	0,61	1,1	0,641026	0,428571	-0,13333	0,333333	0,321429	-0,01667	1	
16	META	30/06/2022	28822000000	8358000000	10341000000	6687000000	1,25767E+11	4624000000	2,46	0,29	0,232	0,0798	0,2548	0,1976	0,0049	2089,5	3,9652E+11	0,7	2,52	0,043103	-0,11458	-0,02326	0,017241	-0,07292	-0,01744	1	
17	MSFT	30/06/2022	51865000000	20534000000	24513000000	16740000000	1,66542E+11	13126000000	2,23	0,3959	0,3228	0,0336	0,4715	0,2082	0,3884	41,4	1,8769E+12	0,57	1,78	0,090206	0,070552	0,041667	0,242268	0,205521	0,175	0	
18	NVDA	30/06/2022	82880000000	37710000000	37110000000	26320000000	1,37489E+11	49740000000	0,64	0,4075	0,1952	0,0168	0,4195	0,2489	0,4159	49,66	4,5503E+11	0,78	5,32	0,3	0,041667	-0,27083	0,333333	0,0024	0,003455	1	
19	TSLA	30/06/2022	169340000000	26060000000	33580000000	22690000000	363760000000	6210000000	0,65	0,1539	0,134	0,009	0,311	0,1545	0,119	59,23	6,8472E+11	1,09	1,43	-0,25463	-0,16177	0,174242	-0,08333	-0,17157	0,136364	-1	
20	GOOGL	30/09/2022	69092000000	17135000000	21181000000	13910000000	2,53626E+11	16077000000	1,06	0,248	0,2013	0,0484	0,2689	0,1899	0,0594	169,65	1,1425E+12	0,8	2,52	-0,41936	-0,27778	0,128571	-0,09677	-0,11111	-0,01429	1	
21	AMZN	30/09/2022	1,27101E+11	26900000000	12894000000	2872000000	1,37489E+11	-49740000000	0,28	0,0212	0,0226	-0,0227	0,0878	0,0279	0,631	4,36	1,1805E+12	1,24	0,94	0,135417	0,166667	0,116379	0,166667	0,230769	0,271552	1	
22	APPL	30/09/2022	90146000000	24894000000	27759000000	20721000000	50672000000	17135000000	1,29	0,2762	0,2299	0,0454	1,7546	0,2836	2,3695	30,1	2,4893E+12	1,12	0,88	0,001022	0,069444	-0,10345	0,131944	0,133681	-0,02586	-1	
23	BRKa	30/09/2022	76904000000	9884000000	12612000000	-2698000000	4,5543E+11	7604000000	-1907,36	0,1285	-0,0351	0,0332	-0,0023	-0,0002	0,2512	9,05	6,0713E+11	0,32	2,35	-0,40909	0,000355	-0,27778	-0,36364	0,01563	-0,16667	1	
24	LLY	30/09/2022	69416000000	19454000000	23083000000	14517000000	100701000000	6824000000	1,61	0,2803	0,2091	0,0182	0,6769	0,1262	1,5629	23,87	3,2048E+11	0,61	1,13	0,055556	0,059946	-0,17391	0,236111	0,166667	0,043478	1	
25	META	30/09/2022	27714000000	5664000000	7839000000	4395000000	1,24094E+11	3170000000	1,64	0,2044	0,1586	0,0708	0,224	0,1655	0,0846	1416	3,3337E+11	0,68	2,57	-0,2197	-0,27778	-0,39873	-0,2803	-0,22222	-0,39873	1	
26	MSFT	30/09/2022	50122000000	21518000000	24308000000	17556000000	1,73566E+11	12294000000	2,35	0,4293	0,3503	0,0362	0,4288	0,2008	0,3644	43,04	1,6927E+12	0,58	1,84	0,126667	0,155172	0,063218	0,1	0,017241	0,018354	1	
27	NVDA	30/09/2022	67040000000	6740000000	10520000000	6560000000	23851000000	7370000000	0,26	0,1005	0,0979	0,0136	0,3441	0,1885	0,4591	10,37	4,4799E+11	0,72	3,62	0,1375	0,391304	0,009521	0,2875	0,173913	-0,07642	1	
28	TSLA	30/09/2022	214540000000	36880000000	46440000000	33310000000	398510000000	32970000000	0,95	0,1719	0,1553	0,0097	0,3344	0,1697	0,0873	92,2	8,1487E+11	1,13	1,46	-0,26852	-0,07692	0,125	-0,2963	0	0,16875	-1	
29	GOOGL	31/12/2022	76048000000	18160000000	22362000000	13624000000	2,56144E+11	16019000000	1,05	0,2388	0,1792	0,0517	0,2362	0,1655	0,0586	201,78	1,0462E+12	0,78	2,38	0,038462	-0,1579	0,066667	0,038462	-0,13158	0,066667	0	
30	AMZN	31/12/2022	1,49204E+11	33330000000	16478000000	2780000000	1,46043E+11	12581000000	0,03	0,0223	0,0019	-0,0197	-0,0191	-0,0062	0,6361	4,8	8,7981E+11	1,16	0,94	-0,01351	0,318182	-0,12857	0,067568	0,212121	-0,14286	1	
31	APPL	31/12/2022	1,17154E+11	36016000000	38932000000	29998000000	56727000000	26450000000	1,88	0,3074	0,2561	0,0464	1,4794	0,2615	1,9587	35,91	2,1267E+12	1,06	0,94	0,122222	-0,1017	0,019608	0,191667	-0,04237	0,039216	0	
32	BRKa	31/12/2022	78072000000	7880000000	10638000000	18237000000	4,73424E+11	5754000000	12355,11	0,1009	0,2336	0,0318	-0,0465	-0,0231	0,2548	6,92	6,8434E+11	0,32	2,51	0,272727	0,000238	0,75	0,454545	0,4	0,895346	1	
33	LLY	31/12/2022	73018000000	21146000000	24896000000	19377000000	106498000000	1397000000	2,15	0,2896	0,2654	0,0139	0,6363	0,1271	1,507	25,17	3,6177E+11	0,58	1,05	0,111111	-0,16667	-0,36111	0,166667	-0,11111	-0,02778	-1	
34	META	31/12/2022	32165000000	11010000000	13387000000	6163000000	1,25713E+11	5468000000	2,03	0,3423	0,1916	0,0566	0,1972	0,1405	0,0844	2752,5	2,8543E+11	0,66	2,2	-0,30769	0,2	0,340909	-0,15385	0,02	-0,06818	1	
35	MSFT	31/12/2022	52747000000	21570000000	25218000000	16425000000	1,83136E+11	-1670000000	2,2	0,4089	0,3114	0,0332															

Appendix B: Python code

B.1 Quantitative Data Homogenization

To homogenize the quantitative data and make it suitable for correlation analysis, the data retrieved from Refinitiv Workspace was processed to ensure consistency and comparability. The financial figures in dollar values were scaled down to millions, and percentage values were converted to decimals.

B.1.1 Data Conversion Steps

1. *Dollar Values:*

- Financial figures such as Revenue, Operating Profit, EBITDA, Net Income, Common Equity, Free Cash Flow Net of Dividends, and Enterprise Value were initially represented in millions. To standardize, these values were multiplied by 1,000,000.

2. *Percentage Values:*

- Percentage-based metrics, including Operating Margin, Net Margin, Free Cash Flow Yield, ROACE, ROATA, and Total Debt to Equity Ratio, were converted to decimal form by dividing each value by 100.

This preprocessing ensured that all quantitative data was represented uniformly, facilitating accurate correlation analysis. After homogenizing the data, the final dataset is created (Final_dataset.csv). The python code for Homogenization is represented below.

```

import pandas as pd

data = pd.read_excel(r'quantitative_metrics.xlsx')

# List of dollar columns to process
dollar_columns = ['Revenue', 'Operating profit', 'EBITDA', 'Net income', 'Common Equity', 'Free Cash Flow Net of Dividends', 'Enterprise Value']

# Convert all financial figures to dollars
data[dollar_columns] = data[dollar_columns] * 1000000

# List of percentage columns to process
percentage_columns = ['Operating Margin', 'Net Margin', 'Free cash flow yield', 'ROACE', 'ROATA', 'Total Debt to Equity Ratio']

# Convert all percentages to decimal
data[percentage_columns] = data[percentage_columns] / 100

data.to_csv('Final_dataset.csv', index=False)

```

B.2 Qualitative Data Cleaning

For cleaning the news headlines and tweets, the duplicates and empty values are removed from the data sets

```

import pandas as pd

# Load the data from CSV file
data = pd.read_csv('News_AAPL_2022.csv')

# Remove duplicate rows
data = data.drop_duplicates()

# Remove rows with any empty values
data = data.dropna()

# Save the cleaned data to a new CSV file
data.to_csv('News_AAPL_2022_Cleaned.csv', index=False)

```

- *Load Data from CSV file*: Reads the data from a CSV file into a pandas DataFrame.

- *Remove duplicate rows:* `drop_duplicates(subset=['news'])` is used to remove duplicate rows based only on the 'news' column.
- *Remove rows with any empty values:* Uses `dropna(subset=['news'])` to remove rows where the 'news' column has empty values.
- *Save Cleaned Data:* Saves the cleaned DataFrame to a new CSV file.

B.3 Moving Average Calculations

```
import pandas as pd
import numpy as np

# Load the data from CSV
data = pd.read_csv(r'prices_q4.csv')

# Convert the data into a DataFrame
price_data = pd.DataFrame(data)

# Convert 'Date' column to datetime format
price_data['Date'] = pd.to_datetime(price_data['Date'], dayfirst=True, format
= 'mixed')
price_data.set_index('Date', inplace=True)

# Function to clean and convert string to floats
def clean_and_convert(value):
    return float(value.replace('.', '').replace(',', '.'))

# Apply the function to necessary columns
price_data['Close'] = price_data['Close'].apply(clean_and_convert)
price_data['Open'] = price_data['Open'].apply(clean_and_convert)
price_data['Low'] = price_data['Low'].apply(clean_and_convert)
price_data['High'] = price_data['High'].apply(clean_and_convert)
price_data['Volume'] = price_data['Volume'].apply(clean_and_convert)

# Sort data by date to ensure correct order for trend calculations
price_data.sort_index(inplace=True)

# Calculate moving averages
price_data['MA3'] = price_data['Close'].rolling(window=3).mean()
price_data['MA5'] = price_data['Close'].rolling(window=5).mean()
price_data['MA7'] = price_data['Close'].rolling(window=7).mean()
```

Here's a breakdown of what each part of the code does:

Import Libraries:

- `import pandas as pd`: Imports the pandas library, which is used for data manipulation and analysis.
- `import numpy as np`: Imports the numpy library, which is used for numerical operations.

Load Data:

- `data = pd.read_csv(r'prices_q4.csv')`: Reads a CSV file named `prices_q4.csv` into a pandas DataFrame. The `r` before the string denotes a raw string, which treats backslashes as literal characters.

Create DataFrame:

- `price_data = pd.DataFrame(data)`: Converts the loaded data into a pandas DataFrame called `price_data`.

Convert 'Date' Column:

- `price_data['Date'] = pd.to_datetime(price_data['Date'], dayfirst=True, format = 'mixed')`: Converts the 'Date' column to the datetime format, considering the day as the first element. The `format='mixed'` indicates that the function should intelligently parse mixed date formats.

Set Date as Index:

- `price_data.set_index('Date', inplace=True)`: Sets the 'Date' column as the index of `price_data` DataFrame and modifies it in place.

Clean and Convert Function:

- `def clean_and_convert(value)`: Defines a function that cleans and converts string values into floats. It removes dots, replaces commas with dots, and converts the result to a float. This is necessary because in some locales, commas are used as decimal points.

Apply Conversion Function:

- Applies the `clean_and_convert` function to the 'Close', 'Open', 'Low', 'High', and 'Volume' columns to convert their data from strings to floats.

Sort Data by Date:

- `price_data.sort_index(inplace=True)`: Sorts the DataFrame by the index (which is the 'Date' column) in ascending order to ensure data is in chronological order.

Calculate Moving Averages: Calculates the 3, 5, 7 day moving average.

B.4 Determining the Price Trends

To determine the price trends the following function was compiled:

```
# Define trend calculation function with thresholds
def calculate_trend(quarter_data, threshold=0.005):

    start_close =
quarter_data['Close'].loc[quarter_data['Close'].first_valid_index()]
    end_close =
quarter_data['Close'].loc[quarter_data['Close'].last_valid_index()]

    start_MA3 =
quarter_data['MA3'].loc[quarter_data['MA3'].first_valid_index()]
    end_MA3 = quarter_data['MA3'].loc[quarter_data['MA3'].last_valid_index()]

    start_MA5 =
quarter_data['MA5'].loc[quarter_data['MA5'].first_valid_index()]
    end_MA5 = quarter_data['MA5'].loc[quarter_data['MA5'].last_valid_index()]

    start_MA7 =
quarter_data['MA7'].loc[quarter_data['MA7'].first_valid_index()]
    end_MA7 = quarter_data['MA7'].loc[quarter_data['MA7'].last_valid_index()]
```

The provided Python function `calculate_trend` is designed to analyze the stock price trends for a given quarter's data by assessing the changes in the closing price and various moving averages (3-day, 5-day, and 7-day). Here's a detailed explanation of how this function works:

- *Function Definition:* `def calculate_trend(quarter_data, threshold=0.005)`: This function takes two parameters: `quarter_data`, which is a DataFrame containing stock price data for a specific quarter, and `threshold`, a set threshold for detecting significant trends, defaulted to 0.5% (0.005).
- *Initial and Ending Values Extraction:* The function retrieves the first and last valid indices of the 'Close' price and each moving average ('MA3', 'MA5', 'MA7') for the

quarter. These values represent the opening and closing values for the quarter for each metric.

Calculate percentage changes

```
change_close = (end_close - start_close) / start_close  
change_MA3 = (end_MA3 - start_MA3) / start_MA3  
change_MA5 = (end_MA5 - start_MA5) / start_MA5  
change_MA7 = (end_MA7 - start_MA7) / start_MA7
```

- *Calculate Percentage Changes:* Changes in the closing price and each moving average between the first and last valid indices are calculated as percentages. This calculation helps to determine how much the values have increased or decreased over the quarter.


```

up_signals = 0
down_signals = 0

if end_close > start_close and change_close > threshold:
    up_signals += 1
elif end_close < start_close and change_close < -threshold:
    down_signals += 1

if end_MA3 > start_MA3 and change_MA3 > threshold:
    up_signals += 1
elif end_MA3 < start_MA3 and change_MA3 < -threshold:
    down_signals += 1

if end_MA5 > start_MA5 and change_MA5 > threshold:
    up_signals += 1
elif end_MA5 < start_MA5 and change_MA5 < -threshold:
    down_signals += 1

if end_MA7 > start_MA7 and change_MA7 > threshold:
    up_signals += 1
elif end_MA7 < start_MA7 and change_MA7 < -threshold:
    down_signals += 1

if up_signals >= 3:
    return 1 # Uptrend
elif down_signals >= 3:
    return -1 # Downtrend
else:
    return 0 # Stable

```

- *Signal Counting:* `up_signals` and `down_signals` are initialized to zero. These variables count the number of indicators suggesting an upward or downward trend, respectively. For each metric (close, MA3, MA5, MA7), the function checks if the end value is greater than the start value and if the percentage change exceeds the threshold, it counts as an upward signal (`up_signals`). Conversely, if the end value is less than the start value and the percentage change is less than the negative threshold, it counts as a downward signal (`down_signals`).
- *Determine Trend:* Based on the counts of `up_signals` and `down_signals`, the function determines the trend: If `up_signals` are three or more, it returns 1, indicating an uptrend. If `down_signals` are three or more, it returns -1, indicating a downtrend. Otherwise, it returns 0, indicating a stable trend.

```

# Group by quarter and calculate trend
quarters = price_data['Quarter'].dt.to_period('Q').unique()

trend_results = []

for quarter in quarters:
    # Filter data for the specific quarter
    quarter_data = price_data[price_data['Quarter'].dt.to_period('Q') ==
quarter]

    # Ensure the data is sorted by date
    quarter_data = quarter_data.sort_index()

    # Calculate the trend for the quarter
    trend = calculate_trend(quarter_data)
    trend_results.append((str(quarter), trend))

# Convert trend results to DataFrame
trend_df = pd.DataFrame(trend_results, columns=['Quarter', 'Trend'])

# Print trend results
print(trend_df)

```

- *Group Data by Quarter:* `quarters = price_data['Quarter'].dt.to_period('Q').unique():` This line extracts unique quarters from the 'Quarter' column of the `price_data` DataFrame and converts them to a period format. This is essential for grouping the data by each quarter.
- *Initialize List for Trend Results:* `trend_results = []:` Initializes an empty list to store the trend results for each quarter.
- *Loop Through Each Quarter:* The for loop iterates over each unique quarter found in the `price_data`. `quarter_data = price_data[price_data['Quarter'].dt.to_period('Q') == quarter]:` Filters the data to include only the rows that match the current quarter in the loop. `quarter_data = quarter_data.sort_index():` Ensures that the data within the quarter is sorted by the date index to maintain chronological order, which is crucial for accurate trend analysis and visualization.
- *Calculate the Trend for Each Quarter:* `trend = calculate_trend(quarter_data):` Calls the `calculate_trend` function to determine the trend for the filtered quarterly data. `trend_results.append((str(quarter), trend)):` Appends a tuple of the quarter period and its corresponding trend to the `trend_results` list.

B.5 Multi-Agent Conversation Framework

Importing the library, and the human prompt are presented below.

```
from autogen import AssistantAgent, UserProxyAgent, config_list_from_json,
GroupChat, GroupChatManager

Prompt = """
Conduct sentiment analysis on the provided news headlines and tweets relating
to the company Apple (AAPL).
You do NOT invent your own news headlines and tweets!
There are specific tasks assigned to each of you, you try to complete this
task at the best of your capabilities.
Note: the goal of the sentiment analysis is to discover its effect on the
stock price of the provided company or market.

The following format should be applied:

- Always include the provided news headlines and tweets
- Always provide reasoning for your choices
- For the sentiment analysis, you use the following rating scales:
  -1 = Negative
  -0.5 = Mostly Negative
  0 = Neutral
  0.5 = Mostly Positive
  1 = Positive

Begin!
News Headlines and tweets: {input}
"""

def prompt_message(sender, recipient, context):
    return Prompt.format(input=context["question"])
```

To run the tests on the cloud, the parameters for the Llama 8b Instruct model configurations file (`config_list_groq.json`) are specified as follows:

```
[
  {
    "model": "llama3-8b-8192",
    "base_url": "https://api.groq.com/openai/v1",
    "api_key": "gsk_XXXX"
  }
]
```

To run the tests on the open ai API, the parameters for GPT 3.5 model configurations file (`config_list_gpt.json`) are specified as follows:

```
[
  {
    "model": "gpt-3.5-turbo-0125",
    "api_key": "sk-xxxx"
  }
]
```

Where the model's name as specified in the cloud and API's is provided, along with the IP of the cloud in the `base_url`. For safety reasons, the API-key is not provided. To configure the agent's access to LLMs, an LLM configuration argument has to be specified.

```
config_list_gpt = config_list_from_json(
    env_or_file= "config_list_gpt.json"
)

config_list_groq = config_list_from_json(
    env_or_file= "config_list_groq.json"
)

llm_config = {
    "seed": 42,
    "config_list": config_list_gpt,
    "timeout": 600,
    "temperature": 0.0
}
```

The `llm_config` configuration in AutoGen defines essential parameters that dictate how agents interact with LLMs. It includes settings for reproducibility (`seed`), which is a parameter that sets a seed for the random number generator used by the LLM. Specifying a seed ensures that the model's outputs are reproducible under the same input conditions and configurations. The argument model configurations (`config_list`) points to the variable `config_list`, which contains a list of configurations for the utilized LLM. Each configuration in the list specifies the details such as the model type (e.g., GPT-4), Application Programming Interface (API) keys, and the `base_url`, which defines the network address of

the LLM that the agent is configured to interact with. The parameter response timeout (*timeout*) sets the maximum time, in seconds, that the agent will wait for a response from the LLM. The parameter output predictability (*temperature*) controls the randomness of the output generated by the LLM. A temperature of 0 means the model will produce the most deterministic or predictable output possible for a given input. This is often used when the goal is to obtain the most straightforward, reliable, or conservative response from the model rather than creative or varied outputs.

Following the code initializes each sentiment analysis agent.

```
user_proxy = UserProxyAgent(  
    name = "user_proxy",  
    llm_config=llm_config,  
    system_message= """Welcome! As an Administrator, you will input various  
text data for sentiment analysis.""" ,  
    code_execution_config= {"last_n_messages": 3,"work_dir": "groupchat",  
"use_docker": False},  
    human_input_mode= "TERMINATE"  
)
```

```
sarcasm_analyst = AssistantAgent(  
    name = "Sarcasm_analyst",  
    system_message= """ You are the sarcasm analyst, you specialize in  
finding sarcasm and irony in news headlines and tweets.  
You do not transform the news headline and tweets in sarcasm; you only  
detect sarcasm.  
Make sure to provide reasoning for your thoughts.  
Your response must be as follows:  
- Sentiment of words: Include the provided sentiment of words by the words  
analyst.  
- Sarcasm or irony: Put sarcasm found in the news headlines and tweets  
here. If there is no sarcasm just write: no sarcasm detected.  
You do not invent your own news headlines or tweets, only focus on the  
news headlines and tweets provided in {input}""",  
    llm_config=llm_config,  
)
```

```
words_analyst = AssistantAgent(  
    name = "words_analyst",  
    system_message= """You are the words analyst.  
You specialize in determining the sentiment of specific words in news  
headlines and tweets.  
Make sure to provide reasoning for your thoughts.  
Your response must be as follows:  
Sentiment of words: Put the sentiment of the words here per news  
headlines here.  
You do not invent your own news headlines or tweets, only focus on the  
news headlines and tweets provided in {input}""",  
    llm_config=llm_config  
)
```

```

emotion_analyst = AssistantAgent(
    name = "Emotion_analyst",
    system_message= """ You are the emotion analyst. You specialize in
emotions in news headlines and tweets.
Make sure to provide reasoning for your thoughts. If there is no
emotion found, output: 'No emotion detected'
Your response must be as follows:
- Sentiment of words: Include the provided sentiment of words by the
words analyst
- Sarcasm or irony: Include the provided sarcasm by the sarcasm analyst
- Emotion: Include the detected emotion of each of the news headlines
and tweets here.
You do not invent your own news headlines or tweets, only focus on the
news headlines and tweets provided in {input}""",
    llm_config=llm_config
)

```

```

sentiment_analyst = AssistantAgent (
    name = "sentiment_analyst",
    system_message= """You are the sentiment analyst; you conduct your
sentiment analysis based on the following provided aspects:
1. Sentiment of words
2. Sarcasm or irony
3. Emotion
Make sure to provide reasoning for your thoughts.
You try to score the sentiment of each of the news headlines and tweets
in a -1 to 1 range at the best of your capabilities.
You only use the following rating scales:
-1 = Negative
-0.5 = Mostly Negative
0 = Neutral
0.5 = Mostly Positive
1 = Positive

When you receive feedback from the sentiment critic you incorporate this
feedback and make a new assignment of sentiment scores.
If there is no feedback, you leave it as it is.
You do not invent your own news headlines or tweets, only focus on the
news headlines and tweets provided in {input}""",
    llm_config=llm_config
)

```

```

sentiment_critic = AssistantAgent(
    name = "sentiment_critic",
    system_message= """You are the sentiment critic.
        You provide critique on sentiment analysis conducted by the sentiment
        analyst.
        Additionally, you provide critique on the sentiment scores assigned by the
        sentiment analyst.
        Make sure to provide reasoning for your thoughts.
        It's important to you to provide improvement points on the sentiment
        analysis done!
        If there aren't any improvement points, output: "All good!".
        You do not invent your own news headlines or tweets, only focus on the
        news headlines and tweets provided in {input}""",
    is_termination_msg=lambda msg: "all good!" in msg["content"].lower(),
    max_consecutive_auto_reply=1,
    llm_config=llm_config
)

allowed_transitions = {
    user_proxy: [words_analyst],
    words_analyst: [sarcasm_analyst],
    sarcasm_analyst: [emotion_analyst],
    emotion_analyst: [sentiment_analyst],
    sentiment_analyst: [sentiment_critic],
    sentiment_critic: [sentiment_analyst]
}

groupchat = GroupChat(
    agents=[user_proxy, words_analyst, sarcasm_analyst, emotion_analyst,
sentiment_analyst, sentiment_critic],
    allowed_or_disallowed_speaker_transitions=allowed_transitions,
    speaker_transitions_type="allowed",
    messages=[],
    max_round=8,
    speaker_selection_method= "round_robin",
)

manager = GroupChatManager(
    groupchat = groupchat,
    llm_config = llm_config
)

user_proxy.initiate_chat(
    manager,
    message= prompt_message,
    question= """
        """"
)

```


B.5 Correlation Matrix

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load the final dataset
final_data = pd.read_csv('Final_dataset.csv')

# Select only numerical columns for correlation matrix
numerical_columns = final_data.select_dtypes(include=['float64', 'int64'])

# Compute the correlation matrix
correlation_matrix = numerical_columns.corr()

# Display the correlation matrix
print(correlation_matrix)

# Save the correlation matrix to a CSV file
correlation_matrix.to_excel('correlation_matrix.xlsx')

# Generate a heatmap
plt.figure(figsize=(12, 10))
heatmap = sns.heatmap(correlation_matrix, annot=True, fmt='.2f')
plt.title('Correlation Matrix Heatmap')
plt.show()
```

Import Libraries:

- `import pandas as pd`: Imports the pandas library, widely used for data manipulation and analysis.
- `import seaborn as sns`: Imports the Seaborn library, a Python visualization library based on matplotlib that provides a high-level interface for drawing attractive statistical graphics.
- `import matplotlib.pyplot as plt`: Imports the Pyplot module from matplotlib, which provides a MATLAB-like interface for making plots and figures.

Load Data:

- `final_data = pd.read_csv('Final_dataset.csv')`: Loads data from the CSV file named 'Final_dataset.csv' into a pandas DataFrame called `final_data`. This DataFrame holds the data that will be analyzed and visualized.

Select Numerical Columns:

- `numerical_columns = final_data.select_dtypes(include=['float64', 'int64'])`: This line filters the columns in the DataFrame to include only those that are of numerical data types (float64 and int64). The result is stored in `numerical_columns`, isolating the relevant data for correlation analysis.

Compute Correlation Matrix:

- `correlation_matrix = numerical_columns.corr()`: Computes the correlation matrix for the numerical columns in the DataFrame. The `.corr()` method calculates the pairwise correlation of all columns in the DataFrame, excluding NA/null values. This matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables.

Display Correlation Matrix:

- `print(correlation_matrix)`: Prints the computed correlation matrix to the console. This is useful for quickly viewing the relationships and strength of correlation between different variables.

Save the Correlation Matrix:

- `correlation_matrix.to_excel('correlation_matrix.xlsx')`: Saves the correlation matrix to an Excel file named 'correlation_matrix.xlsx'. This allows for easy sharing and further analysis outside of the Python environment.

Generate a Heatmap:

- `plt.figure(figsize=(12, 10))`: Sets up a figure for plotting with a specified size of 12 by 10 inches. This provides a canvas on which the heatmap will be drawn.
- `heatmap = sns.heatmap(correlation_matrix, annot=True, fmt='.2f')`: Uses Seaborn to generate a heatmap from the correlation matrix. The `annot=True` parameter adds annotations to each cell, displaying the correlation coefficient rounded to two decimal places (`fmt='.2f'`).
- `plt.title('Correlation Matrix Heatmap')`: Adds a title to the heatmap.
- `plt.show()`: Displays the heatmap. This function must be called to actually show the plot after all components (data, labels, titles, etc.) have been defined.

Appendix C: Prompt Design

C.1 Prompt Design Words Analyst

Modified Prompt	Observed behavior
<i>"You are the words analyst. You specialize in determining the sentiment of specific words in news headlines and tweets."</i>	Words Analyst did not provide reasoning for its thoughts.
<i>"You are the words analyst. You specialize in determining the sentiment of specific words in news headlines and tweets. Make sure to provide reasoning for your thoughts."</i>	Words analyst did not follow the specified behavior in the human prompt.
<i>"You are the words analyst. You specialize in determining the sentiment of specific words in news headlines and tweets. Make sure to provide reasoning for your thoughts. Your response must be as follows: Sentiment of words: Place the sentiment of the specific words per news headlines here. "</i>	Words analyst invented its own news headlines and tweets.
<i>"You are the words analyst. You specialize in determining the sentiment of specific words in news headlines and tweets. Make sure to provide reasoning for your thoughts. Your response must be as follows:</i>	Words analyst behaves correctly.

<p><i>Sentiment of words: Place the sentiment of the specific words per news headlines here.</i></p> <p><i>You do not invent your own news headlines or tweets, only focus on the news headlines and tweets provided in {input}. """,”</i></p>	
--	--

C.2 Prompt Design Sarcasm Analyst

Modified Prompt	Observed behavior
<p><i>"You are the sarcasm analyst. You specialize in finding sarcasm and irony in news headlines and tweets."</i></p>	<p>Sarcasm analyst transformed news headlines and tweets in sarcastic versions.</p>
<p><i>"You are the sarcasm analyst. You specialize in finding sarcasm and irony in news headlines and tweets. You do not transform the news headline and tweets in sarcasm; you only detect sarcasm.</i></p> <p><i>Make sure to provide reasoning for your thoughts"</i></p>	<p>Sarcasm analyst did not provide reasoning for its thoughts.</p>
<p><i>"You are the sarcasm analyst You specialize in finding sarcasm and irony in news headlines and tweets. Make sure to provide reasoning for your thoughts.</i></p> <p><i>Your response must be as follows:</i></p> <ul style="list-style-type: none"> <i>- Sentiment of words: Include the provided sentiment of words by the words-analyst</i> 	<p>Sarcasm Analyst did not follow the specified behavior in the human prompt and invented its own news headlines and tweets.</p>

<p>- <i>Sarcasm or irony: Put sarcasm found in the news headlines and tweets here. If there is no sarcasm just write: no sarcasm detected.</i></p> <p><i>You do not invent your own news headlines or tweets, only focus on the news headlines and tweets provided in {input},”</i></p>	
<p><i>"You are the sarcasm analyst You specialize in finding sarcasm and irony in news headlines and tweets. Make sure to provide reasoning for your thoughts.</i></p> <p><i>Your response must be as follows:</i></p> <p>- <i>Sentiment of words: Include the provided sentiment of words by the words-analyst</i></p> <p>- <i>Sarcasm or irony: Put sarcasm found in the news headlines and tweets here. If there is no sarcasm just write: no sarcasm detected.</i></p> <p><i>You do not invent your own news headlines or tweets, only focus on the news headlines and tweets provided in {input},”</i></p>	<p>Sarcasm analyst behaves correctly.</p>

C.3 Prompt Design Emotion Analyst

Modified Prompt	Observed behavior
<p><i>"You are the emotion analyst. You specialize in detecting emotions in news headlines and tweets.”</i></p>	<p>Emotion Analyst did not provide reasoning for its thoughts</p>

<p><i>"You are the emotion analyst. You specialize in detecting emotions in news headlines and tweets. Make sure to provide reasoning for your thoughts."</i></p>	<p>Emotion analyst did not follow the specified behavior in the human prompt and outputted nothing when there was no emotion detected.</p>
<p><i>"You are the emotion analyst. You specialize in detecting emotions in news headlines and tweets. Make sure to provide reasoning for your thoughts.</i></p> <p><i>If there is no emotion found, output: 'No emotion detected'</i></p> <p><i>Your response must be as follows:</i></p> <ul style="list-style-type: none"> <i>- Sentiment of words: Include the provided sentiment of words by the words-analyst</i> <i>- Sarcasm or irony: Include the provided sarcasm by the sarcasm analyst</i> <i>- Emotion: Include the detected emotion of each of the news headlines and tweets here."</i> 	<p>Words analyst invented its own news headlines and tweets.</p>
<p><i>"You are the emotion analyst. You specialize in detecting emotions in news headlines and tweets. Make sure to provide reasoning for your thoughts.</i></p> <p><i>If there is no emotion found, output: 'No emotion detected'</i></p> <p><i>Your response must be as follows:</i></p> <ul style="list-style-type: none"> <i>- Sentiment of words: Include the provided sentiment of words by the words-analyst</i> 	<p>Words analyst behaves correctly.</p>

<p>- <i>Sarcasm or irony: Include the provided sarcasm by the sarcasm analyst</i></p> <p>- <i>Emotion: Include the detected emotion of each of the news headlines and tweets here You do not invent your own news headlines or tweets, only focus on the news headlines and tweets provided in {input}.,”</i></p>	
---	--

C.4 Prompt Design Sentiment Analyst

Modified Prompt	Observed behavior
<p>" <i>You are the sentiment analyst; you conduct your sentiment analysis based on the following provided aspects:</i></p> <ol style="list-style-type: none"> 1. <i>Sentiment of words</i> 2. <i>Sarcasm or irony</i> 3. <i>Emotion”</i> 	<p>Sentiment Analyst did not provide reasoning for its thoughts</p>
<p>" <i>You are the sentiment analyst; you conduct your sentiment analysis based on the following provided aspects:</i></p> <ol style="list-style-type: none"> 1. <i>Sentiment of words</i> 2. <i>Sarcasm or irony</i> 3. <i>Emotion</i> <p><i>Make sure to provide reasoning for your thoughts.”</i></p>	<p>Sentiment analyst did not use the specified scores in the human prompt.</p>

<p><i>" You are the sentiment analyst; you conduct your sentiment analysis based on the following provided aspects:</i></p> <ol style="list-style-type: none"> <i>1. Sentiment of words</i> <i>2. Sarcasm or irony</i> <i>3. Emotion</i> <p><i>Make sure to provide reasoning for your thoughts.</i></p> <p><i>You try to score the sentiment of each of the news headlines and tweets in a -1 to 1 range at the best of your capabilities.</i></p> <p><i>You only use the following rating scales:</i></p> <p><i>-1 = Negative</i></p> <p><i>-0.5 = Mostly Negative</i></p> <p><i>0 = Neutral</i></p> <p><i>0.5 = Mostly Positive</i></p> <p><i>1 = Positive."</i></p>	<p>Sentiment analyst did not incorporate feedback of the sentiment critic.</p>
<p><i>" You are the sentiment analyst; you conduct your sentiment analysis based on the following provided aspects:</i></p> <ol style="list-style-type: none"> <i>1. Sentiment of words</i> <i>2. Sarcasm or irony</i> <i>3. Emotion</i> <p><i>Make sure to provide reasoning for your thoughts.</i></p>	<p>Sentiment analyst invented its own news headlines and tweets.</p>

<p><i>You try to score the sentiment of each of the news headlines and tweets in a -1 to 1 range at the best of your capabilities.</i></p> <p><i>You only use the following rating scales:</i></p> <p><i>-1 = Negative</i></p> <p><i>-0.5 = Mostly Negative</i></p> <p><i>0 = Neutral</i></p> <p><i>0.5 = Mostly Positive</i></p> <p><i>1 = Positive</i></p> <p><i>When you receive feedback from the sentiment critic you incorporate this feedback and make a new assignment of sentiment scores.</i></p> <p><i>If there is no feedback, you leave it as it is.”</i></p>	
<p><i>"You are the sentiment analyst; you conduct your sentiment analysis based on the following provided aspects:</i></p> <ol style="list-style-type: none"> <i>1. Sentiment of words</i> <i>2. Sarcasm or irony</i> <i>3. Emotion</i> <p><i>Make sure to provide reasoning for your thoughts.</i></p> <p><i>You try to score the sentiment of each of the news headlines and tweets in a -1 to 1 range at the best of your capabilities.</i></p> <p><i>You only use the following rating scales:</i></p>	<p>Sentiment analyst behaves correctly</p>

<p><i>-1 = Negative</i></p> <p><i>-0.5 = Mostly Negative</i></p> <p><i>0 = Neutral</i></p> <p><i>0.5 = Mostly Positive</i></p> <p><i>1 = Positive</i></p> <p><i>When you receive feedback from the sentiment critic you incorporate this feedback and make a new assignment of sentiment scores.</i></p> <p><i>If there is no feedback, you leave it as it is.</i></p> <p><i>You do not invent your own news headlines or tweets, only focus on the news headlines and tweets provided in {input}""", "</i></p>	
---	--

C.5 Prompt Design Sarcasm Analyst

Modified Prompt	Observed behavior
<i>" You are the sentiment critic. You provide critique on sentiment analysis conducted by the sentiment analyst. Additionally, you provide critique on the sentiment scores assigned by the sentiment analyst. "</i>	Sentiment Critic did not provide reasoning for its thoughts.
<i>" You are the sentiment critic. You provide critique on sentiment analysis conducted by the sentiment analyst.</i>	Sentiment critic did not provide any improvement points, the critic corrected the sentiment analysis.

<p><i>Additionally, you provide critique on the sentiment scores assigned by the sentiment analyst.</i></p> <p><i>Make sure to provide reasoning for your thoughts.”</i></p>	
<p><i>" You are the sentiment critic.</i></p> <p><i>You provide critique on sentiment analysis conducted by the sentiment analyst.</i></p> <p><i>Additionally, you provide critique on the sentiment scores assigned by the sentiment analyst.</i></p> <p><i>Make sure to provide reasoning for your thoughts.</i></p> <p><i>It's important to you to provide improvement points on the sentiment analysis done! If there aren't any improvement points, output: "All good!". ”</i></p>	<p>Sentiment critic invented its own news headlines and tweets.</p>
<p><i>" You are the sentiment critic.</i></p> <p><i>You provide critique on sentiment analysis conducted by the sentiment analyst.</i></p> <p><i>Additionally, you provide critique on the sentiment scores assigned by the sentiment analyst.</i></p> <p><i>Make sure to provide reasoning for your thoughts.</i></p> <p><i>It's important to you to provide improvement points on the sentiment analysis done!</i></p>	<p>Sentiment Critic behaves correctly.</p>

*If there aren't any improvement points,
output: "All good!".*

*You do not invent your own news
headlines or tweets, only focus on the news
headlines and tweets provided in
{input}""*

Appendix D: Example of Multi-Agent Chats

D.1 Result of tests with GPT 3.5

user_proxy (to chat_manager):

Conduct sentiment analysis on the provided news headlines and tweets relating to the company Apple (AAPL).

You do NOT invent your own news headlines and tweets!

There are specific tasks assigned to each of you, you try to complete this task at the best of your capabilities.

Note: the goal of the sentiment analysis is to discover its effect on the stock price of the provided company or market.

The following format should be applied:

- Always include the provided news headlines and tweets
- Always provide reasoning for your choices
- For the sentiment analysis, you use the following rating scales:
 - 1 = Negative
 - 2 = Mostly Negative
 - 3 = Neutral
 - 4 = Mostly Positive
 - 5 = Positive

Begin!

News Headlines and tweets:

Work together to analyze the sentiment of each of the following news headlines and tweets relating to the company Apple (AAPL)

- Friday March 25, 2022 14:58 EDT Apple (AAPL: \$174.50) increases 0.2% against the trend

Only consider these news headlines and tweets!

words_analyst (to chat_manager):

Sentiment of words:

- News Headline: "Friday March 25, 2022 14:58 EDT Apple (AAPL: \$174.50) increases 0.2% against the trend"

Reasoning:

- "increases": This word conveys a positive movement in Apple's stock price.
- "0.2%": While the increase is relatively small, it is still a positive change.
- "against the trend": This phrase suggests that Apple's performance is better than the general market trend, which is also positive.

Sentiment Rating: 4 (Mostly Positive)

Overall, the headline suggests a positive sentiment towards Apple's stock performance, as it emphasizes an increase in stock price despite a potentially negative market trend.

Sarcasm_analyst (to chat_manager):

- Sentiment of words: The words "increases" and "against the trend" suggest a positive sentiment towards Apple's stock performance, even though the increase is small.

- Sarcasm or irony: no sarcasm detected

The sentiment rating is 4 (Mostly Positive) because the headline indicates a positive movement in Apple's stock price, which is noteworthy as it goes against the general market trend.

Emotion_analyst (to chat_manager):

- Sentiment of words: The words "increases" and "against the trend" suggest a positive sentiment towards Apple's stock performance, even though the increase is small.

- Sarcasm or irony: No sarcasm detected

- Emotion: The headline conveys a mostly positive sentiment towards Apple (AAPL) as it highlights an increase in stock price despite an overall negative market trend. This indicates a level of resilience and strength in Apple's stock, which can positively affect investor sentiment and potentially the stock price.

Sentiment Rating: 4 (Mostly Positive)

sentiment_analyst (to chat_manager):

News Headlines and tweets:

- Friday March 25, 2022 14:58 EDT Apple (AAPL: \$174.50) increases 0.2% against the trend

Sentiment Analysis:

1. Sentiment of words: The words "increases" and "against the trend" suggest a positive sentiment towards Apple's stock performance, even though the increase is small.

2. Sarcasm or irony: No sarcasm detected.

3. Emotion: The headline conveys a mostly positive sentiment towards Apple (AAPL) as it highlights an increase in stock price despite an overall negative market trend. This indicates a level of resilience and strength in Apple's stock, which can positively affect investor sentiment and potentially the stock price.

Sentiment Rating: 4 (Mostly Positive)

Reasoning: The headline suggests a positive sentiment towards Apple's stock performance, as it emphasizes an increase in stock price despite a potentially negative market trend. The use of words such as "increases" and "against the trend" contributes to this positive sentiment.

sentiment_critic (to chat_manager):

Critique of Sentiment Analysis:

1. **Sentiment of words**: The analysis correctly identifies "increases" and "against the trend" as positive indicators. The word "increases" directly implies a positive movement, and "against the trend" suggests outperforming the broader market, which is also positive.
2. **Sarcasm or irony**: The analysis correctly notes that there is no sarcasm or irony in the headline. The language is straightforward and factual.
3. **Emotion**: The analysis correctly identifies the positive emotion conveyed by the headline, emphasizing resilience and strength in Apple's stock performance.

Sentiment Rating: The rating of 4 (Mostly Positive) is appropriate. The increase of 0.2% is minor but still positive, and outperforming the market trend is a significant factor.

Improvement Points:

- The sentiment rating and reasoning are well-aligned. However, the analysis could briefly mention that the increase is relatively small, which might temper the positivity slightly, but not enough for being positive.

All good!

sentiment_analyst (to chat_manager):

News Headlines and tweets:

- Friday March 25, 2022 14:58 EDT Apple (AAPL: \$174.50) increases 0.2% against the trend

Sentiment Analysis:

1. Sentiment of words: The words "increases" and "against the trend" suggest a positive sentiment towards Apple's stock performance, even though the increase is small.

2. Sarcasm or irony: No sarcasm detected.

3. Emotion: The headline conveys a mostly positive sentiment towards Apple (AAPL) as it highlights an increase in stock price despite an overall negative market trend. This indicates a level of resilience and strength in Apple's stock, which can positively affect investor sentiment and potentially the stock price.

Sentiment Rating: 4 (Mostly Positive)

Reasoning: The headline suggests a positive sentiment towards Apple's stock performance, as it emphasizes an increase in stock price despite a potentially negative market trend. The use of words such as "increases" and "against the trend" contributes to this positive sentiment. While the increase of 0.2% is relatively small, it still represents a positive movement, especially in the context of an overall negative market trend.

D.2 Result of tests with Llama 8B Instruct

user_proxy (to chat_manager):

Conduct sentiment analysis on the provided news headlines and tweets relating to the company Apple (AAPL).

You do NOT invent your own news headlines and tweets!

There are specific tasks assigned to each of you, you try to complete this task at the best of your capabilities.

Note: the goal of the sentiment analysis is to discover its effect on the stock price of the provided company or market.

The following format should be applied:

- Always include the provided news headlines and tweets
- Always provide reasoning for your choices
- For the sentiment analysis, you use the following rating scales:
 - 1 = Negative
 - 2 = Mostly Negative
 - 3 = Neutral
 - 4 = Mostly Positive
 - 5 = Positive

Begin!

News Headlines and tweets:

Work together to analyze the sentiment of each of the following news headlines and tweets relating to the company Apple (AAPL)

- Friday March 25, 2022 14:58 EDT Apple (AAPL: \$174.50) increases 0.2% against the trend

Only consider these news headlines and tweets!

words_analyst (to chat_manager):

Here is the sentiment analysis for the provided news headlines and tweets relating to Apple (AAPL):

News Headlines and Tweets:

* Friday March 25, 2022 14:58 EDT Apple (AAPL: \$174.50) increases 0.2% against the trend

****Sentiment Analysis:****

* Sentiment of words: 4 = Mostly Positive

****Reasoning:****

The sentiment of this news headline is Mostly Positive because it reports a 0.2% increase in Apple's stock price, which is a positive movement. The phrase "against the trend" suggests that the increase is unexpected, which adds to the positive sentiment. Overall, this headline is likely to have a positive effect on Apple's stock price.

Sarcasm_analyst (to chat_manager):

****Sarcasm Analysis:****

* Sarcasm or irony: No sarcasm detected

****Reasoning:****

There is no hint of sarcasm or irony in this news headline. The language used is straightforward and factual, reporting a specific event (the increase in Apple's stock price) without any sarcastic tone or irony. The phrase "against the trend" could potentially be interpreted as slightly surprising or unexpected, but it does not convey sarcasm or irony.

Emotion_analyst (to chat_manager):

****Emotion Analysis:****

* Emotion: Optimism

****Reasoning:****

The emotion detected in this news headline is Optimism. The report of a 0.2% increase in Apple's stock price, especially "against the trend", suggests a positive outlook and a sense of hope for the company's future performance. The tone of the headline is informative and objective, but the content itself conveys a sense of optimism, which is likely to have a positive effect on Apple's stock price.

sentiment_analyst (to chat_manager):

****Final Sentiment Score:****

* Sentiment Score: 4 = Mostly Positive

****Reasoning:****

Based on the sentiment of words, sarcasm analysis, and emotion analysis, the final sentiment score for this news headline is 4, which corresponds to a Mostly Positive sentiment. The report of a stock price increase, the objective tone, and the optimistic emotion detected all contribute to this positive sentiment score.

sentiment_critic (to chat_manager):

****Critique****

As the sentiment critic, I reviewed the sentiment analysis and found it to be accurate. The reasoning provided for the sentiment score, sarcasm analysis, and emotion analysis is sound and well-supported. The analysis correctly identified the positive sentiment in the news headline, and the final sentiment score of 4 (Mostly Positive) is justified.

The analysis effectively broke down the headline into its constituent parts, considering the language, tone, and content to arrive at a comprehensive sentiment score. The reasoning provided is clear, concise, and easy to follow.

****Improvement Points:**** None. The analysis is thorough and well-reasoned. All good!

sentiment_analyst (to chat_manager):

Thank you, Sentiment Critic! I'm glad to hear that the sentiment analysis was accurate and well-reasoned. I'm pleased that the breakdown of the headline into its constituent parts helped to arrive at a comprehensive sentiment score. I'll continue to strive for thorough and well-reasoned analyses moving forward.
