



Nikita Wagner, Mtk. Nr.: 70476751

Experimenteller Vergleich von Stimmungsanalyse und technischer Analyse durch ChatGPT zur Aktienprognose

Abschlussarbeit zur Erlangung des Hochschulgrades

Bachelor of Science

im Studiengang Informatik an der

Ostfalia Hochschule für angewandte Wissenschaften - Hochschule Braunschweig/Wolfenbüttel

Erste/-r Prüfer/-in: Prof. Dr. Wolfgang Ludwig

Zweite/-r Prüfer/-in: Prof. Dr. Denis Royer

Eingereicht am 19.02.2025

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich versichere, dass ich alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet habe, und dass die eingereichte Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen ist.



Wolfenbüttel, den 19. Februar 2025

Zusammenfassung

Die Vorhersage von Aktienkursen ist seit jeher ein zentrales Ziel von Marktteilnehmern, um maximale Profite zu erzielen. Mit der Einführung neuer Technologien, insbesondere Künstlicher Intelligenz, eröffnen sich innovative Ansätze zur effizienten und präzisen Analyse. In der vorliegenden Arbeit wurde das Large Language Model ChatGPT untersucht, das seit seiner Einführung die Nutzung von Künstlicher Intelligenz maßgeblich geprägt hat. Ziel ist es, ChatGPT zur Generierung von Handelsempfehlungen auf Basis zweier Analysemethoden einzusetzen: der technischen Analyse, die ausschließlich auf Preisdaten basiert, und der Stimmungsanalyse, die Nachrichtenartikel und die öffentliche Stimmung zu einer Aktie auswertet. Dabei wurden verschiedene Prompting-Methoden untersucht und in die Prompts integriert, um die Qualität der Ergebnisse zu optimieren. Zudem kamen neue Tools der ChatGPT-API zum Einsatz, um die Untersuchungen weitgehend zu automatisieren und die Vergleichbarkeit der Ergebnisse sicherzustellen.

Um die beiden Analysemethoden zu vergleichen, wurden insgesamt über 3.500 Handelsempfehlungen generiert. Die entwickelte Software ermöglicht die Speicherung der von ChatGPT generierten Handelsempfehlungen in einer Datenbank, einschließlich aller Daten, die das Modell zur Entscheidungsfindung herangezogen hat. Zusätzlich werden die Begründungen für jede Entscheidung gespeichert, um die Nachvollziehbarkeit des Entscheidungsprozesses zu gewährleisten. Im Anschluss wurde eine Handelssimulation durchgeführt, in der separate Handelsstrategien, darunter eine Take-Profit-Strategie und das Schließen aller Positionen am Ende des Handelstages, implementiert wurden.

Die Ergebnisse zeigen, dass beide Ansätze in der Lage waren, positive Renditen zu erzielen. Während die Stimmungsanalyse durch eine höhere Trefferquote überzeugte, erzielte die technische Analyse eine signifikant höhere Rendite. Zudem wurde deutlich, dass die Auswahl der analysierten Unternehmen einen wesentlichen Einfluss auf die Performance der Methoden hatte. Bei der technischen Analyse führten insbesondere große und volatile Unternehmen zu besseren Ergebnissen. Bei der Stimmungsanalyse muss darauf geachtet werden, dass die ausgewählten Unternehmen über ausreichend Medienpräsenz verfügen. Abschließend wird betont, dass die Aussagekraft der Untersuchung durch den Einsatz modernerer Modelle sowie eine Analyse über einen längeren Zeitraum und eine größere Anzahl an Unternehmen weiter verbessert werden könnte.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Gegenstand und Motivation	1
1.2	Problemstellung	2
1.3	Zielsetzung	2
1.4	Aufgabenstellung	2
1.5	Aufgabenpakete	2
2	Grundlagen	4
2.1	Künstliche Intelligenz	4
2.1.1	ChatGPT	6
2.1.2	Prompt Engineering	7
2.2	Aktienmarkt	9
2.2.1	Aktienprognose	10
2.2.2	Technische Analyse	12
2.3	Künstliche Intelligenz in der Aktienmarktanalyse	14
2.3.1	Nutzung von LLMs für die Stimmungsanalyse	16
3	Entwicklung eines Handels- und Analysesystems	18
3.1	Externe APIs	18
3.1.1	Alphavantage	18
3.1.2	ChatGPT API	18
3.1.3	Vergleich von ChatGPT und Alphavantage in Betracht der technischen Analyse	19
3.2	Technischer Aufbau des Handels- und Analysesystems	22
3.2.1	Datenbankstruktur	22
3.2.2	Softwareaufbau und Datenfluss	23
3.3	Vorgehen bei der Simulation	27
3.3.1	Technische Analyse	28
3.3.2	Stimmungsanalyse	29
3.3.3	Realisierung von Gewinnen und Verlusten	30
3.4	Untersuchte Unternehmen	33
3.5	Bewertungskriterien und Zielparameter	34
4	Analyse der Daten	36
4.1	Konsistenz	36
4.2	Effizienz	37

4.3	Praktikabilität	39
4.4	Genaugigkeit.....	40
4.4.1	Gewinn-Verlust-Quote.....	40
4.4.2	Durchschnittliche Rendite pro Trade.....	42
4.4.3	Gesamtrendite	43
4.4.4	Trefferquote	45
4.4.5	Interpretation der Genaugigkeit	46
5	Ergebnis	48
5.1	Technische Analyse.....	48
5.1.1	Schwächen der technischen Analyse	48
5.1.2	Stärken der technischen Analyse	49
5.2	Stimmungsanalyse	49
5.2.1	Schwächen der Stimmungsanalyse	49
5.2.2	Stärken der Stimmungsanalyse	50
5.3	Lessons Learned: Arbeiten mit ChatGPT.....	50
5.3.1	Strukturierte Outputs.....	51
5.3.2	Tool Calling	51
5.3.3	ChatGPT – API	52
6	Zusammenfassung.....	54
7	Diskussion und Ausblick	56
8	Literaturverzeichnis	59
9	Anhang.....	64

Abbildungsverzeichnis

Abbildung 1: Qualität in Verbindung mit dem Preis des LLMs (artificialanalysis.ai).....	6
Abbildung 2: Kursverlauf Rheinmetall Februar - April 2022	11
Abbildung 3: Static Resistance Linie (Chen, 2010).....	13
Abbildung 4: Uptrend Support Linie (Chen, 2010).....	13
Abbildung 5: API-Referenz des Vergleichs von ChatGPT und Alphavantage	19
Abbildung 6: Stimmungsanalyse durch Alphavantage.....	21
Abbildung 7: Stimmungsanalyse durch ChatGPT ohne Artikelinhalt.....	21
Abbildung 8: Stimmungsanalyse durch ChatGPT mit Artikelinhalt	21
Abbildung 9: ERD der Datenbank.....	22
Abbildung 10: Softwareaufbau.....	24
Abbildung 11: Zod Antwortschema.....	25
Abbildung 12: Beispielantwort in gefordertem Format.....	26
Abbildung 13: Beschreibung der durch ChatGPT aufrufbaren Funktionen	27
Abbildung 14: Prozedur zum Auswerten der Handelsempfehlungen.....	32
Abbildung 15A: Vergleich von den Stimmungs- und Relevanzscores von Alphavantage (links) und ChatGPT (rechts)	64
Abbildung 16A: Analyseergebnis von ChatGPT mit Handelsempfehlung und den Begründungen für die Entscheidung.....	64
Abbildung 17A: Standardabweichung AAPL	65
Abbildung 18A: Standardabweichung XOM.....	66
Abbildung 19A: Standardabweichung DKNG	67
Abbildung 20A: Standardabweichung IBIO.....	68
Abbildung 21A: Standardabweichung MRK.....	68
Abbildung 22A: Vergleich von Stimmungsanalyse Alphavantage (links) und ChatGPT (rechts) bei zu vielen Input Token	69

Tabellenverzeichnis

Tabelle 1: Vergleich von Preisdatenanfrage	20
Tabelle 2: Untersuchte Unternehmen nach Sektor, Volatilität und Marktkapitalisierung.....	33
Tabelle 3: Anzahl der Handelsempfehlungen pro Unternehmen, Analysetyp und Handelsempfehlung ...	37
Tabelle 4: Summe und Durchschnitt der verbrauchten Token und Zeit.....	38
Tabelle 5: Summe und Durchschnitt der verwendeten Token und der verbrauchten Zeit von Apple, Microsoft, Tesla und Nvidia	39
Tabelle 6: Gewinn-Verlust-Quote der technischen Analyse und Stimmungsanalyse	40
Tabelle 7: Gewinn-Verlust-Quote gruppiert nach Unternehmensgröße	41
Tabelle 8: Gewinn-Verlust-Quote gruppiert nach Sektor	41
Tabelle 9: Gewinn-Verlust-Quote gruppiert nach Volatilität	42
Tabelle 10: Durchschnittliche Rendite Pro Trade.....	42
Tabelle 11: Gesamtrendite aller Trades	43
Tabelle 12: Gesamtrendite gruppiert nach Volatilität.....	44
Tabelle 13: Gesamtrendite gruppiert nach Sektor	44
Tabelle 14: Trefferquote beider Analysemethoden über alle Unternehmen hinweg	45
Tabelle 15: Trefferquote gruppiert nach Volatilität.....	45
Tabelle 16: Trefferquote gruppiert nach UnternehmensgröÙte	45

1 Einleitung

Im Folgenden wird mit Hilfe der 5 Stufen Methode erklärt in welchem Kontext diese Bachelorarbeit steht und weshalb sie von Relevanz ist, gefolgt von der Problemstellung und dem Ziel dieser Arbeit. Zum Schluss wird die dafür auszuführende Aufgabe in abgegrenzte Arbeitspakete mit überschaubarem Arbeitsaufwand gegliedert.

1.1 Gegenstand und Motivation

Seit der Einführung von OpenAI's Large Language Model (LLM) ChatGPT und dessen kostenfreier Verfügbarkeit hat sich die Nutzung von Künstlicher Intelligenz (KI) in rasantem Tempo verbreitet. Menschen aus den verschiedensten Bereichen greifen durch die niedrigschwellige Verfügbarkeit auf ChatGPT für eine Vielzahl von Anwendungsfällen zurück [1] wie beispielsweise das Auffinden von Kochrezepten, die Beantwortung von Gesundheitsfragen, die Unterstützung bei Hausaufgaben oder Programmierproblemen, um nur einige zu nennen. Für viele stellt dieses Modell den ersten Schritt in die Welt der KI dar. Es bieten sich jedoch nicht nur für private Nutzer Vorteile, sondern auch Unternehmen profitieren von der Verbreitung dieser Technologie. Der Finanzbereich war einer der ersten, der künstliche Intelligenz nutzte, um beispielsweise Trends anhand von Finanzdaten oder auch Nachrichten zu erkennen und so ihre Effizienz zu steigern [2].

Trotz seiner Vielseitigkeit und Leistungsfähigkeit stößt ChatGPT jedoch regelmäßig an seine Grenzen: Es liefert gelegentlich fehlerhafte oder unvollständige Antworten, da es auf ungenaue oder unzuverlässige Daten zurückgreift oder vorsichtige Aussagen vermeidet [3]. Diese Unzuverlässigkeit stellt die Nutzer vor die Herausforderung, präzise Anweisungen zu formulieren, um qualitativ hochwertige und verlässliche Antworten zu erhalten. Die Beziehung zwischen der Qualität der Eingaben (Prompts) und den resultierenden Ausgaben eines LLMs ist dabei äußerst komplex. Unterschiedliche Strategien wie Few-Shot Learning und Reasoning werden eingesetzt, um die Qualität der Anfragen zu optimieren. [4] Doch selbst mit ausgeklügelten Prompt-Strategien bleibt eine gewisse Unsicherheit bestehen, da das Modell möglicherweise auf fehlerhafte Grundinformationen zurückgreift [5]. Es erscheint daher sinnvoll, das Modell durch zusätzliche, verlässlich recherchierte Informationen zu unterstützen, um die Genauigkeit der Antworten zu steigern.

Denn vor allem im Bereich der Finanzmärkte ist eine hohe Genauigkeit essenziell, weshalb es klar definierte Kennzahlen zur Bewertung von Unternehmen gibt, wie das Kurs-Gewinn-Verhältnis (KGV), das Kurs-Buchwert-Verhältnis (KBV), Relative Stärke Index (RSI) oder die Dividendenrendite. Diese Kennzahlen spiegeln die Gesundheit und Profitabilität eines Unternehmens wider und geben Aufschluss darüber, ob ein Unternehmen unter- oder überbewertet ist. [6] Doch nicht nur Finanzkennzahlen beeinflussen den Finanzmarkt, auch Nachrichten haben erheblichen Einfluss auf die Aktienkurse. Investoren, die in Aktien investieren, verfolgen die Nachrichtenlage, um Informationen über den Zustand ihrer Unternehmen zu erhalten. Studien belegen, dass sich sowohl positiv als auch negativ konnotierte Nachrichten im Aktienkurs widerspiegeln. [7]

In dieser Bachelorarbeit wird untersucht, ob ChatGPT bessere Ergebnisse im Aktienmarkt erzielt, wenn es Nachrichten analysiert, d. h. eine Stimmungsanalyse durchführt, oder wenn es Finanzdaten analysiert, also eine technische Analyse vornimmt. Zudem wird untersucht, welche Daten ChatGPT benötigt, um eine gute Leistung zu erzielen, und ob durch reines Prompt Engineering ohne zusätzliche Daten

belastbare Annahmen getroffen werden können. Es gibt bereits einige Untersuchungen der Leistung von Stimmungsanalyse und der technischen Analyse. Es gibt jedoch keinen hinreichend guten Vergleich zwischen den beiden Analysen. Außerdem hat keine Untersuchung die neue Funktion der ChatGPT-API genutzt, in der strukturierte Ausgaben zu einer verlässlichen Schnittstelle führen könnten. Das ist relevant, da es Auswertungen gibt, in denen eine strukturierte Aussage zu einer schlechteren Leistung des LLMs führt [8].

Diese Bachelorarbeit leistet einen wertvollen Beitrag zur Optimierung der Nutzung von LLMs wie ChatGPT in der Finanzprognose, insbesondere angesichts ihrer wachsenden Popularität und Bedeutung für die Zukunft. Bislang gibt es nur wenige reale Anwendungsfälle im Finanzsektor, sodass dieses Experiment auch als Impuls dienen könnte. Die Ergebnisse dieser Arbeit könnten Anwendern helfen, präzisere und relevantere Anfragen zu formulieren und somit die Qualität der erhaltenen Antworten signifikant zu verbessern. Dies ist von besonderer Relevanz in einer Welt, in der zuverlässige Daten und präzise Analysen immer wichtiger werden, insbesondere im Bereich der Finanzmärkte, wo fehlerhafte Informationen schwerwiegende wirtschaftliche Folgen haben können.

1.2 Problemstellung

- P1) Ein direkter Vergleich zwischen Aktienprognosen mittels Stimmungsanalyse und der technischen Analyse durch ChatGPT ist bisher nicht vorhanden.

1.3 Zielsetzung

- Z1) Die Vor- und Nachteile der Stimmungsanalyse und der technischen Analyse durch ChatGPT sind deutlich erkennbar.

1.4 Aufgabenstellung

- A1) Erstellung einer Software, welche Finanzkennzahlen und Nachrichten für ausgewählte US - Unternehmen aus einer Datenbank ausliest und mit verschiedenen vorgegebenen Prompts an die ChatGPT-API sendet.
- A2) Auswertung der Handelsentscheidung von ChatGPT durch simulierte Käufe und Verkäufe, sowie das Speichern der erzielten Rendite.
- A2) Untersuchung und Dokumentation der Erkenntnisse aus der Nutzung von ChatGPT, einschließlich der Analyse von Erfolgsfaktoren, Herausforderungen und Optimierungsmöglichkeiten.

1.5 Aufgabenpakete

- AP1.1) Untersuchung von möglichen Optimierungsmöglichkeiten durch Prompt Engineering.
- AP1.2) Auswahl und Begründung der zu untersuchenden Unternehmen.

- AP1.3) Planung und Erstellung der für die Leistungsanalyse benötigten Tabellen in einer relationalen Datenbank und speichern historischer Finanzdaten und Nachrichten.
 - AP1.4) Implementierung eines NodeJS-Backends, welches in der Lage ist den Namen einer Aktie zu lesen und für diese Aktie von ChatGPT eine Prognose zu erhalten.
- AP2.1) Speichern der Rendite jeder ausgewählten Aktie in einem bestimmten Zeitraum für die ausgewählte Methode.
 - AP2.2) Analyse und Visualisierung der durchschnittlichen Renditen der verschiedenen Methoden.

2 Grundlagen

Im Folgenden werden die Grundlagen vorgestellt, die für die Einordnung dieser Arbeit relevant sind. Elementar hierbei ist Künstliche Intelligenz, mit einem besonderen Fokus auf LLMs, da diese in den weiteren Kapiteln thematisiert und untersucht werden. Außerdem wird Prompt Engineering als Teil der Künstlichen Intelligenz vorgestellt. Danach wird erläutert, was der Aktienhandel ist, welche Kennzahlen in dem Zusammenhang relevant sind und wie versucht wird, Aktienprognosen zu erstellen. Zuletzt wird erklärt, welchen Platz Künstliche Intelligenz im Finanzbereich hat, aber insbesondere bei der Aktienprognose.

2.1 Künstliche Intelligenz

Künstliche Intelligenz (KI) ist ein Teilgebiet der Informatik, das sich mit der Entwicklung von Systemen befasst, die in der Lage sind, Aufgaben auszuführen, die üblicherweise menschliche Intelligenz erfordern. Dabei umfasst KI die Fähigkeit, aus Daten zu lernen, Muster zu erkennen, Entscheidungen zu treffen oder Probleme zu lösen. Eine der zentralen Methoden zur Realisierung solcher Systeme ist das maschinelle Lernen. [9]

Beim maschinellen Lernen werden Algorithmen mit Trainingsdaten trainiert, die häufig aus Paaren von Eingaben und den zugehörigen Ausgaben bestehen. Beispielsweise kann ein Algorithmus lernen, E-Mails als „Spam“ oder „kein Spam“ zu klassifizieren, indem er eine Vielzahl solcher gekennzeichneten Daten analysiert. Ziel des Trainings ist es, dass der Algorithmus allgemeine Muster und Zusammenhänge erkennt, um später auch bei neuen, unbekannten Daten korrekte Vorhersagen treffen zu können. Dieser Ansatz wird als überwachtes Lernen (Supervised Learning) bezeichnet und ist eine der grundlegenden Methoden innerhalb des maschinellen Lernens. [10]

Anders als klassische maschinelle Lernmodelle, die meist für klar definierte Aufgaben trainiert werden, basiert die Entwicklung von Large Language Models (LLMs) auf einem skalierbaren und umfassenderen Trainingsansatz. [11]

Large Language Models sind Sprachmodelle, die auf neuronalen Netzen basieren und mit einer großen Vielfalt an Textdaten trainiert werden. Im Unterschied zu klassischen Algorithmen, die mit spezifischen Eingabe-Ausgabe-Paaren trainiert werden, erfolgt das Training von LLMs auf der Grundlage unstrukturierter und umfangreicher Textsammlungen, die aus verschiedenen digitalen Quellen wie Büchern, wissenschaftlichen Artikeln, Nachrichten, Webseiten und sozialen Medien stammen. Ziel dieses Trainings ist es, ein breites Verständnis von Sprache, Kontext und Bedeutungen zu entwickeln. [9,12]

Der Trainingsprozess von LLMs basiert auf dem Prinzip des unüberwachten Lernens. Dabei wird das Modell darauf trainiert, die nächste Wortfolge oder Textsequenz auf Basis eines gegebenen Kontexts vorherzusagen. Ein Beispiel: Wenn der Eingabetext lautet „Das Wetter ist heute“, berechnet das Modell die wahrscheinlichste Fortsetzung, beispielsweise „sonnig“ oder „regnerisch“. Dieser Ansatz erfordert keine explizit gekennzeichneten Trainingsdaten, sondern ermöglicht es dem Modell, durch die Analyse von Mustern in den Textdaten selbstständig Strukturen und Zusammenhänge zu erlernen. Dadurch erhält es eine noch neutralere Wissensbasis, als wenn man die Daten kennzeichnet. [11,13]

Ein wesentlicher Unterschied zum Training klassischer Algorithmen besteht darin, dass LLMs nicht auf eine einzelne spezifische Aufgabe ausgelegt sind. Während ein Spam-Filter ausschließlich darauf trainiert wird, E-Mails zu klassifizieren, sind LLMs in der Lage, flexibel auf eine Vielzahl von Aufgaben zu reagieren. Diese Flexibilität wird durch die enorm große Menge an Parametern ermöglicht, die das Modell während des Trainingsprozesses optimiert. [14,15] Dieses Merkmal von LLMs ist das sogenannte „In-Context Learning“: Diese Modelle können aus dem Kontext einer Eingabe Anforderungen ableiten und die wahrscheinlichste Antwort generieren, ohne explizit für diese spezifische Aufgabe trainiert worden zu sein. Diese Eigenschaft hebt LLMs von klassischen, streng spezialisierten Modellen ab und ermöglicht eine hohe Flexibilität in der Anwendung. [15,16]

Nach dem initialen Training auf allgemeinen Textdaten können LLMs durch einen weiteren Trainingsschritt, das sogenannte Fine-Tuning, auf spezifische Anwendungsbereiche spezialisiert werden. Beim Fine-Tuning werden dem Modell domänenspezifische Daten zur Verfügung gestellt, wie beispielsweise medizinische Fallberichte, juristische Texte oder technische Dokumentationen. In diesem Prozess kommen häufig wieder Eingabe-Ausgabe-Paare zum Einsatz, ähnlich wie beim überwachten Lernen. In der Literatur wird dieses Vorgehen „semi-supervised learning“ genannt. Ein Modell, das beispielsweise in der Medizin eingesetzt werden soll, könnte mit Patientenbeschreibungen (Eingabe) und Diagnosen (Ausgabe) weiter trainiert werden, um spezifische Fachfragen präziser beantworten zu können [2,12,13]. Es muss dabei aber beachtet werden, dass durch das Verändern der Gewichtung in den Trainingsdaten durch Fine-Tuning, das generelle Verständnis der Sprache und Allgemeinwissen verloren gehen kann, weshalb eine Balance gefunden werden muss, die je nach Nutzen unterschiedlich gewichtet ist. [11] Untersuchungen haben ergeben, dass Fine-Tuned LLMs, wenn überhaupt, nur einen kleinen Vorteil zu generellen LLMs haben [2], dafür aber keine intensive Trainingsphase brauchen, sondern „out-of-the-box“ gute Ergebnisse liefern. Dazu müssen jedoch die richtigen Anfragen an die LLM formuliert werden, welche im Kapitel Prompt Engineering vorgestellt und diskutiert werden.

Dieser zweistufige Ansatz – ein breites Training auf allgemeinen Textdaten, gefolgt von einer Spezialisierung durch Fine-Tuning – erlaubt es, Modelle zu entwickeln, die sowohl eine breite Wissensbasis bewahren, als auch in spezifischen Fachbereichen hohe Genauigkeit bieten.

Eine der größten Herausforderungen bei der Entwicklung von LLMs ist die Sicherstellung einer ausgewogenen und qualitativ hochwertigen Datenbasis. Verzerrungen (Biases) können entstehen, wenn bestimmte Perspektiven, Themen oder Gruppen in den Trainingsdaten überrepräsentiert oder unterrepräsentiert sind. Solche Verzerrungen können zu unerwünschten Ergebnissen führen, insbesondere in sensiblen Anwendungsbereichen wie der Medizin, der Rechtswissenschaft oder der Finanzwelt. [17,18] Zurzeit gibt es für dieses Problem aber keinen universellen Standard für die Bewertung von Verzerrungen in den Daten, sodass sich das Messen der Verzerrung schwierig gestaltet [19].

LLMs sind sehr gut darin mit vielen Daten zu arbeiten und diese zu verarbeiten, um zum Beispiel Text Mining zu machen; also Texte auf Informationen zu einem bestimmten Thema zu untersuchen. [2] Durch die vielen Parameter im Training kann ChatGPT 4 Tests in verschiedenen Bereichen wie Recht, Medizin [15] und Finanzbuchhaltung [20] bestehen, was beweist, dass sein Grundwissen ausreichend für auch sehr anspruchsvolle Domänen ist. Die Texte, die dabei entstehen sind kaum von Menschen geschriebenen Texten zu unterscheiden [18] und haben im Vergleich zu Machine Learning den Vorteil, dass sie menschenähnliche Begründungen für ihre Entscheidungen liefern können [21].

Obwohl LLMs sehr vielversprechend sind, haben sie dennoch Schwächen. Sie sind schlecht darin numerische Aufgaben zu lösen, in denen Rechnung gemacht werden müssen, um Entscheidungen zu

treffen [11]. Eben wurde erwähnt, dass ChatGPT 4 Tests zu Finanzbuchhaltung besteht. Solche Tests bestehen aber zum Großteil aus Fragen die Antworten erwarten, die keine Rechnung oder Erklärung enthalten, was das Bewerten von LLMs für den Finanzbereich verkompliziert und die Kompetenz in Frage stellt [2]. Und obwohl LLMs menschenähnliche Logik anwenden können und diese erklären können, sind menschliche Analysen, die von Experten in dem Bereich durchgeführt wurden, deutlich detaillierter [21], was wiederum sehr wichtig ist, wenn mit den Antworten des Modells weitergearbeitet wird [22]. Forscher von Apple [23] haben die Hypothese aufgestellt, dass LLMs eigentlich nicht echte Erklärungen liefern können, zumindest bei mathematischen Fragen, sondern nur in ihren vorhandenen Daten Erklärungen suchen, die zu der Situation passen. Das habe die Konsequenz, dass sobald eine Frage etwas anders formuliert ist oder irrelevante Informationen enthält, sich das Ergebnis signifikant verschlechtert.

2.1.1 ChatGPT

In der Untersuchung wird das Modell GPT-4o Mini verwendet. Die Gründe für diese Wahl werden im Folgenden dargelegt.

Die Internetplattform artificialanalysis.ai vergleicht verschiedene Large Language Models in unterschiedlichen Kategorien miteinander, darunter die Qualität der generierten Antworten, wissenschaftliches Wissen, quantitative Fähigkeiten (z. B. Mathematik), Programmierungsfähigkeiten und weitere Aspekte. In diesen Tests erzielte OpenAi's Modell o1 durchweg die besten Ergebnisse und übertraf alle anderen Modelle, unabhängig vom Hersteller. Auch o1-mini belegte in den meisten Aufgaben den zweiten Platz. Googles Modell Gemini 1.5 Pro sowie Anthropic's Claude 3.5 schnitten in den Tests ebenfalls gut ab und lagen in ihrer Leistung nahe beieinander. GPT-4o Mini erreichte in allen getesteten Kategorien eine geringere Leistung als GPT-4o und befindet sich auf einem vergleichbaren Niveau mit Gemini 1.5 Flash. Die Open-Source-Modelle LLaMA 3.1 von Meta mit 8 Milliarden Parametern erzielten hingegen in allen Tests die schletesten Ergebnisse. [24]

Quality vs. Price

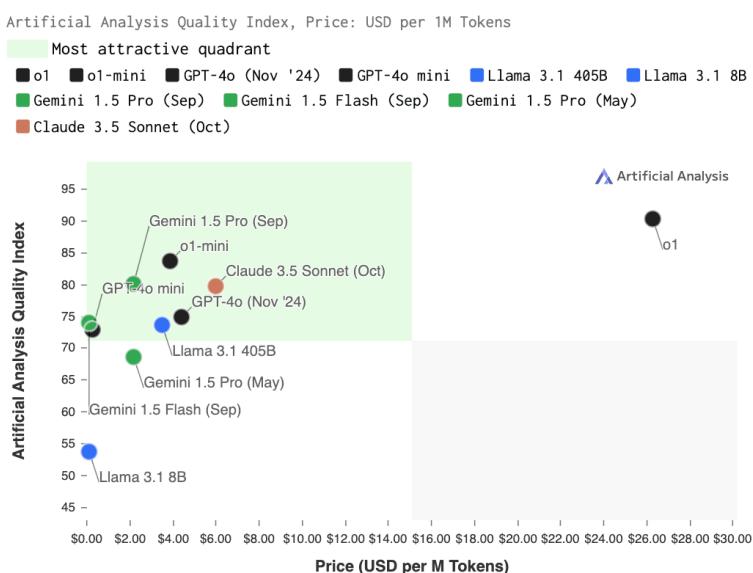


Abbildung 1: Qualität in Verbindung mit dem Preis des LLMs (artificialanalysis.ai)

Ein Blick auf die Kosten zeigt, dass Gemini 1.5 Flash nur die Hälfte der Kosten von GPT-4o Mini verursacht, mit einem Preis von 0,07 USD pro Million Input-Token und 0,3 USD pro Million Output-Token. Auch DeepSeek V2 ist günstiger als GPT-4o Mini. Das Modell o1 ist 100-mal teurer als GPT-4o Mini, während GPT-4o bereits 17-mal teurer ist. Googles Pro-Modelle sind zwar günstiger als o1 und GPT-4o, jedoch mit 5 USD pro Million Output-Token immer noch vergleichsweise teuer [24]. Vor diesem Hintergrund stellt sich die Frage, warum in der vorliegenden Untersuchung GPT-4o Mini verwendet wurde und nicht Gemini 1.5 Flash, das ähnliche Funktionen bietet, die für diese Untersuchung relevant sind, und zudem eine geringere Latenz aufweist. Ein Vergleich der Modelle in Abbildung 1 zeigt, dass Gemini 1.5 Flash hinsichtlich der Kosten am weitesten links positioniert ist, mit GPT-4o Mini also zu den günstigsten Modellen gehört, und sich auf einer Ebene mit LLaMA 3.1 (8B) befindet. Diese Ähnlichkeit würde Gemini 1.5 Flash zu einer nahezu identischen Alternative für den vorliegenden Anwendungsfall machen [24].

Die Entscheidung für GPT-4o Mini lässt sich jedoch durch die Marktverbreitung und Akzeptanz von ChatGPT begründen. Laut eines Berichts der Internetseite First Page Sage [25] entfallen 59,5% des Marktanteils von KI-gestützten Chatbots auf ChatGPT, gefolgt von Microsoft Copilot mit 14,3% und Google Gemini mit 13,4%. Die Frankfurter Allgemeine Zeitung berichtet [26], dass ChatGPT im Juni 2024 2,9 Milliarden Besucher verzeichnete, während Google Gemini lediglich 400 Millionen monatliche Nutzer hatte. Diese Zahlen verdeutlichen die weitaus größere Verbreitung von ChatGPT im Vergleich zu anderen Modellen.

Humlum und Vestergaard [27] heben hervor, dass bereits ein Jahr nach der Einführung von ChatGPT 80% der Software Entwickler in Dänemark diese Technologie nutzen, während 34% der Finanzberater ebenfalls auf diese Lösung zurückgreifen. Zudem ist ChatGPT aufgrund seiner frühen Markteinführung und seiner hohen Funktionalität einem breiten Publikum bekannt. Dies hat dazu geführt, dass viele Menschen das Modell ausprobieren und es sich als Standard etabliert hat. Brynjolfsson et al. [28] betonen, dass ChatGPT insbesondere für unerfahrene Arbeitskräfte eine hohe Effektivität aufweist. Diese breite Akzeptanz und Verbreitung von ChatGPT und seinen Varianten, wie GPT-4o Mini, macht es zu einer naheliegenden Wahl für die vorliegende Untersuchung, da es wahrscheinlich von den meisten Menschen als erstes in Betracht gezogen wird.

Unabhängig davon, welches LLM verwendet wird, ist die richtige Formulierung von Anfragen entscheidend für die Qualität der generierten Ergebnisse. Im folgenden Kapitel wird daher das Konzept des Prompt Engineering vorgestellt, das sich mit Techniken zur effektiven Gestaltung von Prompts an LLMs beschäftigt.

2.1.2 Prompt Engineering

Prompt Engineering ist eine Schlüsselstrategie, um Large Language Models gezielt einzusetzen und ihre Leistung zu optimieren. Die Qualität eines Prompts hat nachweislich einen erheblichen Einfluss auf die generierten Antworten [4]. Durch präzise Gestaltung können Modelle relevantere und besser abgestimmte Ergebnisse liefern. Dabei sind die Definition von Zielen, Vorgaben zu Format und Länge sowie eine klare Sprache zentral für den Erfolg [29].

Zero-Shot Prompting ist die grundlegendste Methode, bei der ein Modell ohne vorherige Beispiele oder spezifischen Kontext angewiesen wird, eine Aufgabe zu lösen. Ein typisches Beispiel wäre die Aufforderung:

„Was ist die Hauptstadt von Deutschland?“

Die LLM wird die Stadt nennen, aber das Format und der Detailgrad der Ergebnisse, werden durch diese Art von Anfragen nicht spezifiziert und können daher auch nicht vorhergesagt werden. Diese Methode erfordert keine Vorbereitung und ist daher effizient sowie vielseitig einsetzbar [14]. Schwächen zeigen sich jedoch bei komplexeren Aufgaben, da fehlender Kontext zu unpräzisen oder fehlerhaften Ergebnissen führen kann [4]. Es wird empfohlen, Aufgaben präzise zu formulieren und das gewünschte Ergebnis, einschließlich Format und Länge, klar anzugeben, um Missverständnisse zu vermeiden [29]. Für komplizierte Szenarien ist die Zerlegung der Aufgabe in kleinere, sequenzielle Schritte vorteilhaft, da dies die Fehleranfälligkeit reduziert [4].

Few-Shot Prompting erweitert die Eingabe durch mehrere Beispiele, die dem Modell helfen, den gewünschten Kontext und die Aufgabenstruktur zu verstehen, wie zum Beispiel:

„Beantworte folgende Fragen:“

1. *Was ist die Hauptstadt von Italien? Rom.*
2. *Was ist die Hauptstadt von Frankreich? Paris.*
3. *Was ist die Hauptstadt von Deutschland?“*

Es wurde gezeigt, dass Modelle durch diese Methode auch ohne zusätzliches Finetuning präzisere Ergebnisse liefern können, die mit spezialisierten Modellen vergleichbar sind [11]. Besonders bei kreativen oder analytischen Aufgaben verbessert die Bereitstellung hochwertiger Beispiele die Genauigkeit und Konsistenz erheblich [2]. In einem Experiment konnte eine Accounting-Prüfung durch die Verwendung von 10-Beispiel-Prompts um 6,6 % bessere Ergebnisse erzielen [20]. Nachteile ergeben sich jedoch aus der höheren Token Anzahl, die längere Rechenzeiten und gesteigerte Kosten verursachen kann. Token sind die „Bausteine“ aus denen Wörter im Kontext von LLMs bestehen. Ein Token entspricht im Durchschnitt 4 englischen Zeichen [30].

Das **Chain-of-Thought (CoT) Prompting** fordert das Modell auf, seine Denkprozesse in mehreren Schritten explizit darzulegen, bevor eine endgültige Antwort gegeben wird. Diese Methode hat sich insbesondere bei mehrschichtigen oder logisch anspruchsvollen Aufgaben als vorteilhaft erwiesen. Ein einfaches Beispiel dafür sieht wie folgt aus:

„Beantworte folgende Frage und erkläre mir Schritt für Schritt, wie du zu dieser Lösung gekommen bist: Was ist die Hauptstadt von Deutschland?“

In einer Studie zur Finanzanalyse konnte durch die Anwendung von CoT die Genauigkeit der Antworten von 52 % auf 60 % gesteigert werden, indem das Modell schrittweise zum Ziel geführt wurde [21]. Die Zerlegung komplexer Aufgaben in kleinere, modularisierte Einheiten minimiert die Fehlerquote und erleichtert die Verarbeitung [4]. Zudem wird durch klare Zwischenschritte die Transparenz und Nachvollziehbarkeit der Antwort erhöht, während ein strukturierter Prompt das Modell bei der Bearbeitung unterstützt [29]. Kojima et al. [31] zeigten, dass bei GPT-3 mit CoT Prompting Ergebnisse ähnlich zu den Verbesserungen durch Few Shot Learning sind und einfacher umzusetzen sind, da keine Beispiele gefunden werden müssen, sondern nur „Let’s think step by step.“ in der Prompt stehen muss. Mit einer zusätzlichen Abfrage wurde dann aus der ersten Antwort das Ergebnis ohne Erklärung herausgefiltert.

Instruction Prompting gibt dem Modell detaillierte Anweisungen, einschließlich des gewünschten Kontexts, des Formats und des Stils. Ein typisches Beispiel wäre:

„Du bist Topografieexperte. Lies die Frage sorgfältig und antworte nur mit einem Wort. Was ist die Hauptstadt von Deutschland?“

Diese Methode fördert präzisere und relevantere Antworten, da die Erwartungen klar definiert werden [32]. Eine präzise und klar formulierte Sprache, ergänzt durch deutliche Trennzeichen wie Anführungszeichen, reduziert die Fehlerrate und verbessert die Ergebnisse [29]. Zudem wird empfohlen, beide Seiten eines Themas zu beleuchten, um Verzerrungen in der Antwort zu vermeiden [32]. Der Aufwand für die Erstellung solcher Prompts ist jedoch höher, und die Tokenanzahl steigt mit der Länge der Anweisungen.

Self-Consistency Prompting erweitert Chain-of-Thought, indem das Modell aufgefordert wird, mehrere Antworten unabhängig voneinander zu generieren. Die beste Antwort wird anschließend ausgewählt.

Diese Technik erhöht die Zuverlässigkeit und Präzision der Ergebnisse, insbesondere bei Wissensabfragen [33]. Iterative Ansätze und die explizite Vorgabe mehrerer möglicher Outputs können die Verarbeitung komplexer Aufgaben unterstützen, während konsistente Antworten Fehler minimieren [29]. Der höhere Rechenaufwand bleibt jedoch ein bedeutender Nachteil.

Die Wahl der richtigen Prompting-Strategie beeinflusst maßgeblich die Effektivität und Präzision von LLMs. Zero-Shot- und Few-Shot-Methoden eignen sich für einfache oder grundlegende Aufgaben, während CoT- und Self-Consistency-Techniken komplexe Szenarien bewältigen können. Präzise Anweisungen, eine klare Aufgabenstruktur und die Zerlegung in kleinere Schritte tragen zur Minimierung von Fehlern und zur Optimierung der Ergebnisse bei [4,29]. Kombinierte Ansätze und eine kontinuierliche Verfeinerung der Prompts können die Potenziale von LLMs weiter ausschöpfen.

2.2 Aktienmarkt

Um die Mechanismen des Aktienhandels sowie die Analyse und Prognose von Kursentwicklungen zu verstehen, ist eine solide theoretische Grundlage erforderlich. Dieses Kapitel erläutert zunächst die wesentlichen Merkmale von Unternehmen, die nach dem US-amerikanischen Recht als Corporations firmieren, sowie deren Aktien, bevor es auf den Börsenhandel und die Abgrenzung der betrachteten Finanzprodukte eingeht. Im Rahmen dieser Arbeit liegt der Fokus explizit auf Aktien von Corporations. Andere Finanzprodukte wie Anleihen, Rohstoffe oder Investmentfonds bleiben bewusst unberücksichtigt, da Aktien auf dem Finanzmarkt meistens die erste Anlaufstelle für interessierte Einsteiger sind.

Eine Corporation ist nach dem Recht der Vereinigten Staaten eine juristische Person, die unabhängig von ihren Anteilseignern, den Shareholdern, handeln kann. Das Eigenkapital einer Corporation wird in Form von Aktien (Shares) aufgebracht, die das Recht auf eine Beteiligung am Unternehmenserfolg sowie bestimmte Mitspracherechte gewähren. Das Grundkapital (Capital Stock) einer Corporation ist in der Gründungsurkunde (Articles of Incorporation) festgelegt, die bei der zuständigen Bundesstaatenbehörde hinterlegt wird. Aktien verbrieften Eigentumsrechte und können dem Inhaber entweder Stimmrechte bei der Hauptversammlung (Shareholders' Meeting), wo sie zum Beispiel Vorstandsmitglieder wählen können oder eine bevorzugte Behandlung bei der Dividendenzahlung garantieren [34].

Ein zentraler Vorteil dieser Gesellschaftsform liegt in der Haftungsbegrenzung: Aktionäre haften ausschließlich mit ihrer Einlage, während ihr Privatvermögen von der Haftung ausgeschlossen ist. Dieses Prinzip der Limited Liability hat wesentlich zur Attraktivität von Corporations als Organisationsform in den USA beigetragen [35].

Die in einer Corporation ausgegebenen Aktien lassen sich nach verschiedenen Kriterien kategorisieren. Der wesentliche Unterschied besteht zwischen Common Stock und Preferred Stock. Common Stock, vergleichbar mit deutschen Stammaktien, gewährt den Aktionären ein Stimmrecht bei der Hauptversammlung sowie das Recht auf Gewinnbeteiligung, das jedoch nachrangig ist. Preferred Stock hingegen beinhaltet häufig keine Stimmrechte, bietet jedoch vorrangige Dividendenzahlungen. Diese Struktur ermöglicht es Unternehmen, verschiedene Aktionärsgruppen mit individuellen Interessen anzusprechen [36].

Ein weiterer Unterschied betrifft die Übertragbarkeit von Aktien. Die meisten in den USA ausgegebenen Aktien sind als Bearer Shares ausgestaltet, was bedeutet, dass sie frei handelbar sind und keine Registrierung des Eigentümers erfordern. Hingegen erfordern Registered Shares, die insbesondere von kleineren Corporations bevorzugt werden, eine Eintragung des Eigentümers ins Aktienregister, um Eigentumsverhältnisse transparent zu halten und rechtliche Ansprüche nachzuverfolgen zu können [35].

Nicht alle Aktien sind an öffentlichen Börsen handelbar. Viele Unternehmen entscheiden sich bewusst dafür, Aktien im Rahmen privater Platzierungen (Private Placements) zu emittieren, was vor allem in den frühen Entwicklungsphasen einer Corporation von Vorteil ist. Öffentlich gehandelte Aktien (Publicly Traded Shares) werden hingegen an organisierten Märkten wie der New York Stock Exchange (NYSE) oder der NASDAQ gehandelt. Diese Börsen dienen als Plattformen, um Angebot und Nachfrage zusammenzuführen und Marktpreise zu bestimmen. Die Liquidität – also die Möglichkeit, Aktien schnell und zu einem fairen Marktwert zu handeln – ist ein zentrales Merkmal dieser Märkte und wird durch Marktteilnehmer wie Market Maker oder elektronische Handelssysteme unterstützt [34].

Der Börsenhandel unterliegt in den USA einer strikten Regulierung durch die Securities and Exchange Commission (SEC). Diese Behörde überwacht den Handel und stellt sicher, dass Transparenz gewährleistet ist und Marktmisbrauch sowie Manipulation verhindert werden. Insbesondere Gesetze wie der Securities Act von 1933 und der Securities Exchange Act von 1934 bilden die Grundlage für diese Regulierung und garantieren den Schutz von Investoren [37]

Es ist wichtig, den Aktienhandel von anderen Finanzprodukten zu unterscheiden, die ebenfalls an US-amerikanischen Börsen gehandelt werden können. Dazu zählen beispielsweise Corporate Bonds, die von Unternehmen als Schuldtitle ausgegeben werden, sowie Derivate wie Optionen oder Futures, deren Wert an einen Basiswert wie Aktien oder Rohstoffe gekoppelt ist. Rohstoffe wie Öl oder Edelmetalle werden ebenfalls über spezialisierte Märkte wie die Chicago Mercantile Exchange (CME) gehandelt. Diese Finanzinstrumente unterscheiden sich jedoch grundlegend von Aktien, da sie keine Eigentumsrechte an Unternehmen repräsentieren.

2.2.1 Aktienprognose

Bereits im 18. Jahrhundert wurde versucht, den Preisverlauf eines Finanzprodukts vorherzusagen, um strategisch kluge Handelsentscheidungen zu treffen. [38,39] Das Wort „vorhersagen“ ist in diesem Kontext eigentlich nicht besonders passend, da die exakte Vorhersage von Aktienpreisen grundsätzlich nicht möglich ist, da die Preisentwicklung durch eine Vielzahl von Faktoren beeinflusst wird, darunter

zufällige und nicht vorhersehbare Ereignisse, welche zum „Random Walk Pattern“ führen [7,40]. Stattdessen zielt die Analyse darauf ab, mittels statistischer Methoden und verfügbaren Informationen Modelle zu erstellen, die Wahrscheinlichkeiten für bestimmte Preisentwicklungen quantifizieren. Diese Modelle basieren auf historischen Daten, Markttrends, ökonomischen Indikatoren und weiteren relevanten Variablen, um Muster und Zusammenhänge zu erkennen, die zukünftige Preisbewegungen mit einer gewissen statistischen Wahrscheinlichkeit beschreiben können.

Die zwei verbreitetsten Methoden den Finanzmarkt zu analysieren und auf Basis dessen Entscheidungen zu treffen, sind die **Fundamentalanalyse** und die **technische Analyse**. In der vorliegenden Untersuchung wird jedoch nur ein Teilbereich der Fundamentalanalyse untersucht, der **Stimmungsanalyse** genannt wird. Finanzanalysten und Fondmanager nutzen meistens eine Kombination aus beiden Methoden [41] es gibt jedoch auch Trader, die sich ausschließlich auf eine Methode fokussieren [38]. Eigentlich ist eine Kombination aus beiden Methoden nicht vereinbar, wenn man sich die Basen der Analyseansätze anschaut.

Die technische Analyse ist in einer Zeit entstanden, in der es noch nicht verpflichtend war für Unternehmen ihre Finanzkennzahlen an die Öffentlichkeit zu bringen, weshalb man als Basis für den Kauf nur Nachrichten und den Preisverlauf der Aktie kannte [39]. Es gibt die „Efficient Market Hypothesis“, welche besagt, dass die Preise von Wertpapieren in einem effizienten Markt immer im Preis widergespiegelt sind [42]. Das bedeutet, dass es keine asymmetrische Information gibt, also niemand einen Vorteil hat, da alle Informationen, die vorhanden sind, bereits im Preis enthalten sind. Auf dieser Basis argumentieren viele technische Analysten und daher stammt das Argument, dass man sich nur den Preisverlauf anschauen muss, da Daten, wie beispielsweise der Umsatz eines Unternehmens keinen Vorteil bringen.



Abbildung 2: Kursverlauf Rheinmetall Februar - April 2022

Allerdings zeigen Untersuchungen von Lo [43], sowieso Timmermann und Granger [44], dass diese Hypothese nicht immer zutrifft, da irrationale Entscheidungen am Markt existieren und auch asymmetrische Informationsverteilung besteht, was bedeutet, dass einige Marktteilnehmer zu einem Zeitpunkt über mehr Wissen verfügen als andere. Dieses Phänomen lässt sich zum Beispiel am Rüstungsunternehmen Rheinmetall beobachten. Nachdem der russische Präsident Putin am 24.02.2022 der Ukraine den Krieg erklärt hat, ist der Preis der Rheinmetall Aktie von ungefähr 97 Euro auf ca. 107 Euro gestiegen. Eine Woche nach der Ankündigung war der Preis bereits bei über 150 Euro (Abbildung 2). Man sieht also, dass Marktteilnehmer, die früher und besser informiert waren, bei einem niedrigeren Preis eingestiegen sind und mehr Profit erzielen. Doch nicht nur Neuigkeiten sollen laut Analysten Einfluss auf den Preis haben, sondern auch andere Kennzahlen, wie zum Beispiel Umsatz, Gewinn,

Dividend Yield und noch einige mehr [22,45,46]. Fundamental Analysten argumentieren deshalb, dass nicht alle Informationen bereits im Preis enthalten sind und versuchen dies auszunutzen.

Finanzanalysten beginnen im Analyseprozess der Fundamentalanalyse mit der Betrachtung allgemeiner Informationen, wie beispielsweise der Marktkapitalisierung, dem Umsatzwachstum oder aktuellen Nachrichten, und vertiefen die Untersuchung sukzessive. Dabei analysieren sie detailliertere Aspekte wie die Bilanzstruktur, Wettbewerbsvorteile und weitere relevante Faktoren. Eine vollständige Analyse innerhalb einer untersuchten Kategorie umfasst in der Regel zwischen 1.000 und 7.800 Wörter und braucht in der Untersuchung von Bouwman et al. [46] zwischen 19 und 80 Minuten, wodurch eine präzise und fundierte Beschreibung des Unternehmens und seiner Daten gewährleistet wird. Dies bildet die Grundlage für eine qualifizierte Entscheidung hinsichtlich eines Kaufs oder Verkaufs.

Die vorliegende Untersuchung konzentriert sich wie erwähnt ausschließlich auf den spezifischen Teilbereich der Fundamentalanalyse, der als Stimmungsanalyse bezeichnet wird. Die Stimmungsanalyse zielt darauf ab, die allgemeine Marktstimmung zu erfassen, indem untersucht wird, ob aktuell vorwiegend positive Nachrichten vorliegen oder ob negative Neuigkeiten existieren, die potenziell zu panikartigen Reaktionen der Investoren führen könnten. Wenn man als erstes die wahrscheinlich anstehende Stimmung der Massen kennt und auf Basis dessen zuerst eine Aktie kauft, hat man einen besseren Einstiegspunkt als jemand, der die Nachricht erst 2 Tage später erfährt (siehe Rheinmetall Beispiel) und hat somit einen größeren Profit. Nachrichten haben einen signifikanten Einfluss auf den Aktienmarkt, da sie die Wahrnehmung und Bewertung von Unternehmen maßgeblich beeinflussen [47].

Auf der anderen Seite wird untersucht, inwieweit das LLM in der Lage ist, technische Analysen effizient und präzise durchzuführen. Da dieser Themenbereich nicht besonders bekannt ist und in dem Experiment dem LLM zum Beispiel das Moving Average Crossover als beispielhafte Trading Strategie vorgeschlagen wird, wird im folgenden Kapitel die technische Analyse näher vorgestellt.

2.2.2 Technische Analyse

Die technische Analyse ist ein seit langem umstrittenes Thema, da die Effektivität nicht eindeutig wissenschaftlich bewiesen werden kann [48]. Es wird eher von einer selbsterfüllenden Prophezeiung gesprochen, da die „Kursvorhersagen“ der Analysen, Erwartungen sind, die sich allein dadurch bewahrheiten, dass sie getroffen wurden und entsprechendes Verhalten hervorrufen. Das heißt, dass nur dadurch, dass eine Vielzahl an Menschen und Algorithmen so handeln, wie es die technische Analyse erwartet, auch die Kursänderungen oder Trends so passieren, wie es die technische Analyse erwartet. Dafür wird das Massenverhalten der Aktionäre genutzt. [38,48] Die technische Analyse stützt sich auf die Annahme, dass die Märkte durch Emotionen der Massen beeinflusst werden, was sich im Preisverhalten niederschlägt.

Man könnte also auch sagen, dass die technische Analyse nicht versucht den Preis vorherzusagen, sondern das Verhalten der Menschen. Das ist an einfachen Beispielen ablesbar:

Viele Menschen sind bereit dazu, an einem lokalen Hochpunkt des Kurses ihre Anteile zu verkaufen, da sie der Meinung sind, dass der Preis hoch genug sei, um Gewinne zu realisieren. An diesem Hochpunkt gibt es Menschen, die der Meinung sind, dass ein guter Zeitpunkt ist, Leerverkäufe zu starten - auch „Short Selling“ (shorten) genannt. Dabei leihst man dem Markt Aktien zu dem aktuellen Preis in der Hoffnung, dass der Preis der Aktie sinkt und man sie für einen tieferen Preis zurückkaufen kann. Die Differenz der Preise ist der Profit vor Steuern. Dadurch, dass es Menschen gibt, die ihre Aktien verkaufen

und andere, die shorten, entsteht ein sogenannter Verkaufsdruck. Wenn dieser Verkaufsdruck höher ist als der Druck der Käufer, dann sinkt der Preis und die selbsterfüllende Prophezeiung - also, dass der Kurs sich an lokalen Hochpunkten dreht – geht in Erfüllung. Das gleiche soll auch bei lokalen Tiefpunkten funktionieren. [38]

Wenn mehrere lokale Extrema hintereinander den gleichen Preis haben, kann man an den Hochpunkten eine sogenannte statische „Resistance“ Linie (Abbildung 3) ziehen und an den Tiefpunkten eine statische „Support“ Linie. Das ist auch die Basis für die simpelste Handelsstrategie in der technischen Analyse. Es wird erwartet, dass der Preis an diesen Linien abprallt. Das bedeutet, dass viele Marktteilnehmer einen „Long Trade“ eingehen, also eine Aktie kaufen, wenn der Preis die Support Linie erreicht, weil sie erwarten, dass der Preistrend sich an der Linie wendet und wieder steigt. Das führt wieder zu einem Preisdruck, der den Preis oft steigen lässt. Wenn der Preis aber nicht seinen Kurs ändert, ist es ein Indiz dafür, dass der Trend stark ist und der Preis sich wahrscheinlich noch weiter in diese Richtung bewegt. Das gleiche Prinzip kann man auch bei dynamischen Support- und Resistance Linien anwenden. Diese formen sich aus lokalen Hoch- und Tiefpunkten, die aber nicht horizontal den gleichen Preis haben, sondern der Anstieg oder Abstieg ihrer Preisdifferenz linear ist (Abbildung 4).

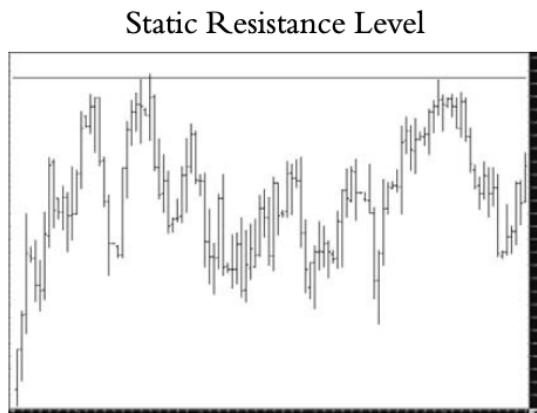


Abbildung 3: Static Resistance Linie (Chen, 2010)

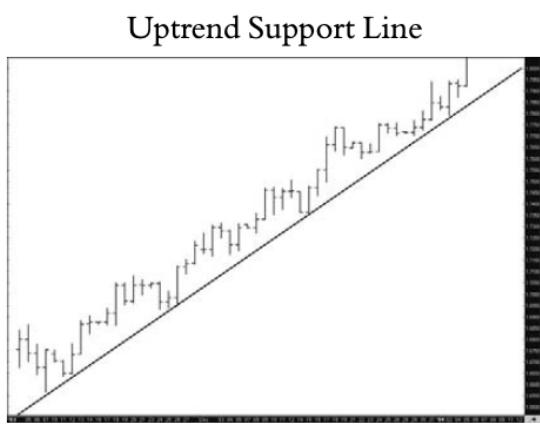


Abbildung 4: Uptrend Support Linie (Chen, 2010)

Eine weitere Strategie, die schon sehr lange genutzt wird, ist das Handeln mit Moving Averages auch „Moving Averages Crossover Strategie“ genannt. [38,39,49,50] Moving Averages stellen gleitende Durchschnittswerte eines Preises über einen definierten Zeitraum dar. Mit zunehmender Länge des betrachteten Zeitraums reagieren diese Durchschnittswerte langsamer auf kurzfristige Schwankungen,

wodurch sie Rauschen effektiver glätten und den übergeordneten Trend klarer sichtbar machen. Kürzere Zeiträume hingegen sind empfindlicher gegenüber kurzfristigen Veränderungen und eignen sich zur Identifikation schneller Trendwechsel. [38]

In Studien von Brock et al. [39] und Sullivan et al [49], die verschiedene Moving-Average-Crossover-Strategien untersuchten, ergab sich in beiden Fällen eine Genauigkeit von über 50 %. Dies deutet darauf hin, dass diese Strategien effektive Ergebnisse liefern können.

Die Moving-Average-Crossover-Strategie arbeitet mit zwei gleitenden Durchschnitten, einem für einen kürzeren Zeitraum (K) und einem für einen längeren Zeitraum (L). Ein Kaufsignal entsteht, wenn der kürzere gleitende Durchschnitt von unten nach oben den längeren Durchschnitt kreuzt, während ein Verkaufssignal generiert wird, sobald K den L von oben nach unten schneidet. Zusätzlich kann die Strategie genutzt werden, um Short-Positionen zu eröffnen: Wenn K unter L fällt, wird eine Short-Position eingegangen, die wieder geschlossen wird, sobald K erneut über L steigt. Die Schnittpunkte der beiden Linien dienen somit als wichtige Indikatoren für den Ein- und Ausstieg in Marktpositionen. [38,39,49]

In der Welt der technischen Analyse existiert eine Vielzahl unterschiedlicher Strategien, wie beispielsweise Strategien basierend auf dem Relative Strength Index, dem Swing Index, dem Triple Screen Trading System oder der Fibonacci-Analyse, um nur einige zu nennen. Eine detaillierte Betrachtung all dieser Ansätze würde jedoch den Rahmen dieser Arbeit übersteigen. [38,51]

In der Regel nutzen erfahrene Trader beim Handeln mit Aktien bewährte Strategien, wie beispielsweise Moving Average Crossovers, die sie konsequent anwenden. Das zentrale Problem dabei ist jedoch, dass spezifische Strategien häufig nur in bestimmten Marktphasen besonders effektiv sind. Wenn eine Marktphase eintritt, in der die gewählte Strategie nicht optimal funktioniert, fehlt möglicherweise die Fähigkeit, die für diese Phase am besten geeignete Vorgehensweise zu identifizieren. Künstliche Intelligenz bietet hier einen Vorteil, da sie den Markt analysieren und die aus ihrer Sicht optimale Strategie für die jeweilige Situation auswählen könnte, was insbesondere für Anfänger von großem Nutzen wäre. Außerdem wird der Markt immer effizienter [49]. Es werden schneller Handelsentscheidungen durch automatisierte Ansätze getroffen, weshalb computergestützte Systeme immer relevanter werden.

Ein weiterer zentraler Aspekt der technischen Analyse ist die Festlegung von Stop-Loss- und Take-Profit-Marken. Der Stop-Loss-Wert definiert den prozentualen Verlust, bei dessen Erreichen eine Position automatisch geschlossen wird, um größere Verluste zu vermeiden. Im Gegensatz dazu legt der Take-Profit-Wert das Gewinnziel fest, ab dem eine Position geschlossen wird, um realisierte Gewinne zu sichern. Die systematische Anwendung dieser beiden Werte stellt eine etablierte Methode in der technischen Analyse dar, um das Risiko- und Ertragsmanagement zu optimieren. [38]

2.3 Künstliche Intelligenz in der Aktienmarktanalyse

Die Anwendung von Large Language Models im Finanzmarkt umfasst zahlreiche Bereiche, die von Banken und Versicherungen bis hin zu Kreditbewertungen und Kundensupport reichen. [2] Innerhalb dieses breiten Spektrums soll der Fokus in diesem Kapitel auf den Einsatz von LLMs in der Aktienmarktanalyse gerichtet werden. Hier liegt der zentrale Nutzen in der Vorhersage von

Kursbewegungen, der Optimierung von Portfolios und der Stimmungsanalyse, da diese Aufgaben für den Handel und die Entscheidungsfindung an den Finanzmärkten von essenzieller Bedeutung sind. [2]

Menschen sind wenig effizient darin, große Datenmengen in kurzer Zeit zu verarbeiten und diese unmittelbar im Gedächtnis zu behalten. Für einen Finanzanalysten ist dies jedoch unerlässlich, weshalb sie aufgrund ihrer Erfahrung oft mit vorgefertigten Szenarien arbeiten, die es ermöglichen Unternehmen in spezifische Kategorien einzuordnen. Diese Vorgehensweise ist äußerst effizient, da die Kategorisierung anhand bestimmter Merkmale schneller und weniger ressourcenintensiv ist als eine vollständige, immer wiederkehrende Neubewertung. [46]

Daher kamen bereits früh klassische Machine-Learning-Ansätze zur Anwendung, insbesondere für die Schätzung von Gewinnen und Verlusten. Die Weiterentwicklung hin zu Deep-Learning-Modellen wie Long Short-Term Memory (LSTM)-Netzwerken oder Transformers führte jedoch zu signifikanten Fortschritten in der Prognosegenauigkeit. LSTM ist eine spezielle Art von wiederkehrendem neuronalen Netzwerk (Recurrent Neural Network, RNN), das darauf ausgelegt ist, Sequenzen von Daten zu analysieren und dabei wichtige Informationen über lange Zeiträume hinweg zu speichern [52,53]. Transformers hingegen sind Modelle, die durch ihre Fähigkeit, große Mengen an Daten parallel zu verarbeiten und dabei Abhängigkeiten zwischen weit entfernten Datenpunkten zu erkennen, besonders effizient sind. [54] LLMs bauen auf diesen Technologien auf und bieten zusätzliche Möglichkeiten, da sie sowohl strukturierte Daten wie historische Preise als auch unstrukturierte Informationen wie Nachrichtenartikel und Social Media Beiträge verarbeiten können. Dies erweitert die Entscheidungsgrundlage für den Handel erheblich [2,7].

Um dem in Kapitel 2.2.1 erklärten Random Walk Pattern entgegenzuwirken, kombinieren moderne Ansätze historische Kursdaten mit Textdaten, um die Prognosekraft zu steigern. Beispielsweise zeigte eine Studie von Lee et al. [15], dass GPT-4 trotz Zero-Shot-Prompting eine Genauigkeit von 54 % bei der Vorhersage von Kursbewegungen erzielen konnte. Dabei wurde darauf hingewiesen, dass zusätzliche finanzielle Metriken zur Bewertung der Ergebnisse erforderlich sind, um valide Aussagen zu ermöglichen, da alleinige Kursdaten und einfache Schwellenwerte oft nicht ausreichen [15]. Neben der Kursvorhersage spielt auch die Stimmungsanalyse eine entscheidende Rolle. LLMs ermöglichen durch Text Mining die Analyse von Nachrichtenquellen, um die allgemeine Haltung oder Emotionen in Bezug auf bestimmte Aktien zu beurteilen. Diese Methode übertrifft durch das LLM oft traditionelle Ansätze der Stimmungsanalyse, die lediglich auf der Häufigkeit bestimmter Schlüsselwörter basieren, da LLMs die Fähigkeit besitzen, den Kontext und die Feinheiten von Texten zu verstehen [55]. Der bloße Umstand, dass ein Satz weniger positiv konnotierte Wörter enthält, bedeutet nämlich nicht zwangsläufig, dass sein Kontext negativ geprägt ist.

Insbesondere bei kleineren Unternehmen, die weniger von automatisierten Handelsstrategien dominiert werden, hat sich gezeigt, dass der Einsatz von LLMs effektiv sein kann, da sie schnellere und präzisere Einschätzungen ermöglichen. [21]

Auch bei der Portfoliooptimierung zeigen LLMs vielversprechende Ansätze. So kann ein trainiertes Modell wie Auto-GPT die Risikobereitschaft eines Nutzers erfassen, relevante Finanzdaten analysieren und mathematische Optimierungen wie die Berechnung der Sharpe Ratio durchführen. Die Sharpe Ratio ist eine Kennzahl, die das Verhältnis von Rendite zu Volatilität bewertet und dabei hilft, das Risiko einer Investition im Vergleich zu ihrer erwarteten Rendite zu verstehen. Während früher für jeden Schritt dieses Prozesses spezialisierte Modelle erforderlich waren, ermöglicht eine LLM die Integration all dieser Aufgaben in einem einzigen Modell, was die Effizienz erheblich steigert [2]. In der Praxis sind

proprietäre Modelle wie GPT-4 oder Gemini Pro oft leistungsfähiger als Open-Source-Alternativen, was die Wahl geeigneter Technologien beeinflusst [21].

Trotz der vielversprechenden Ergebnisse stehen LLMs vor spezifischen Herausforderungen. Eine davon ist die Schwierigkeit, finanzspezifische Aufgaben durchzuführen, die sowohl quantitative als auch qualitative Fähigkeiten erfordern. Während LLMs gut darin sind, große Datenmengen zu analysieren und Muster zu erkennen, fehlt es ihnen oft an der Fähigkeit, spezialisierte Finanzkenntnisse direkt aus ihrem Modellgedächtnis abzuleiten. Dies erfordert eine detaillierte Aufbereitung der Eingabedaten und maßgeschneiderte Prompts, um valide Ergebnisse zu erzielen [21]. Hinzu kommt, dass der Wettbewerbsvorteil durch den Einsatz von LLMs sinken kann, je mehr Marktteilnehmer auf ähnliche Technologien zurückgreifen, da der Markt in der Lage ist, schnell auf neue Informationen zu reagieren und diese einzupreisen [55]. Zusätzlich können unzureichend konfigurierte oder falsch trainierte Modelle zu Fehlentscheidungen führen, was in volatilen Märkten zu erheblichen Verlusten führen kann. Trotz dieser Herausforderungen zeigt sich, dass der Einsatz von LLMs im Aktienhandel das Potenzial hat, traditionelle Ansätze zu ergänzen und in bestimmten Bereichen zu übertreffen. Dies macht sie zu einem wertvollen Werkzeug in der modernen Finanzwelt.

2.3.1 Nutzung von LLMs für die Stimmungsanalyse

Die Untersuchung der Literatur zu der Nutzung von LLMs für die technische Analyse und Stimmungsanalyse offenbart diverse Ansätze. Ein Großteil der Methoden basiert auf der Kategorisierung von Ergebnissen in verschiedene – zumeist drei – Cluster: Anstieg, Rückgang oder keine Veränderung. Ziel ist es, dass das LLM anhand unterschiedlicher Datensätze bestimmt, welchem dieser Cluster der Preisverlauf einer Aktie zugeordnet werden kann [7,22]. Einige Untersuchungen erweitern diese Analyse durch die Bewertung der Stärke eines Trends, beispielsweise durch die Unterscheidung zwischen einem starken Anstieg und einem schwachen Rückgang [21].

Für die Stimmungsanalyse existieren verschiedene Informationsquellen. Einige Ansätze verwenden Twitter-Daten [7,40,56], während andere auf Nachrichtenüberschriften [57], vollständige Nachrichtenartikel [58,59] oder Ausschnitte aus Jahresberichten von Unternehmen [22,60] zurückgreifen. Bei der Nutzung dieser Quellen ist es von entscheidender Bedeutung, ihre Relevanz für das analysierte Unternehmen zu prüfen, bevor sie bewertet werden [57]. So ist beispielsweise ein positiver Nachrichtenartikel über ein neues Modell eines Automobilherstellers wenig aussagekräftig, wenn das Unternehmen sich gleichzeitig in einer wirtschaftlichen Krise befindet. Die Relevanzprüfung kann entweder manuell erfolgen, durch eine automatisierte Entscheidung des LLM [57] oder durch die Berücksichtigung eines Relevanzwertes des Nachrichtenanbieters, wie etwa von Alphavantage.

Untersuchungen haben ergeben, dass Nachrichten ihren größten Einfluss kurz nach Veröffentlichung haben [55,58]. Lira und Tang [55] entwickelten ein Programm, das auf Basis von Nachrichten automatisiert Aktienkäufe und -verkäufe durchführt. Besonders erfolgreich war dabei ein Modell, das Aktien 15 Minuten nach Veröffentlichung einer Nachricht erwarb und am Ende desselben Tages wieder veräußerte.

Qiu et al. [22] untersuchten den Einfluss von Jahresberichten mithilfe von SVM-basierten Modellen. Dabei war es erforderlich, eine Trainingsgruppe zu erstellen, in der spezifische Wörter gefiltert und angepasst wurden, um eine übermäßige Assoziation bestimmter Begriffe mit den Ergebnissen zu vermeiden. Diese Vorarbeit entfällt bei der Verwendung von LLMs, da diese den Kontext und die

Stimmung auch aus komplexen Texten analysieren können [55]. Die Analysen zeigten, dass kurzfristige Aktienkursveränderungen erfolgreich anhand von 10-K-Filings prognostiziert werden konnten.

Insbesondere wurde festgestellt, dass Jahresberichte in klarer Sprache mit positiven Kursentwicklungen korrelierten, während schlechte Unternehmensleistungen häufig durch Berichte mit umfangreichem Fachjargon und indirekten Formulierungen schwerer zu identifizieren waren [22,60].

Insgesamt bewerteten die Studien die Ansätze durchweg positiv, wobei Vorhersagewahrscheinlichkeiten über 50 % erzielt wurden. Ob vergleichbare Ergebnisse in dieser Untersuchung mithilfe von ChatGPT erzielt werden können, wird in den folgenden Kapiteln dargelegt.

3 Entwicklung eines Handels- und Analysesystems

Um die technische Analyse mit der Stimmungsanalyse unter Einsatz eines LLM objektiv vergleichen zu können, wird ein System benötigt, das eine standardisierte und reproduzierbare Vergleichsbasis schafft. Im Folgenden werden der Aufbau dieses Systems, die zugrundeliegenden Komponenten sowie die Methodik des Vergleichs der beiden Ansätze erläutert.

3.1 Externe APIs

ChatGPT ist vielen Menschen als interaktiver Chatbot über die Plattform *chatgpt.com* bekannt. Darüber hinaus besteht jedoch die Möglichkeit, API-Anfragen an die [ChatGPT-API](#) zu senden, was sich als äußerst praktisch für die Entwicklung automatisierter Systeme erweist.

Für das Sammeln historischer Marktdaten stehen mehrere Optionen zur Verfügung, wie beispielsweise *Yahoo Finance* oder *Google Finance*. Da es jedoch mühsam ist, sämtliche Daten bereitzustellen und deren Qualität zu gewährleisten, wird der lizenzierte Partner der NASDAQ, [Alphavantage](#), genutzt. Es wird Alphavantage genutzt, da sie sehr viele Finanzdaten kostenlos über ihre API zur Verfügung stellen und gleichzeitig auch Nachrichtendaten kostenlos anbieten. Bei anderen bekannten Anbietern ist dieser Aspekt mit monatlichen Kosten verknüpft. Um eine größere Vielfalt an Nachrichtenartikeln bereitzustellen, wird ergänzend die Plattform [gnews.io](#) genutzt, da die kostenlose Nutzung ausreichend viele API-Aufrufe ermöglicht und sehr einfach zu implementieren ist.

3.1.1 Alphavantage

Alphavantage bietet eine kostenlose Version ihrer API an, die bis zu 25 Anfragen pro Tag ermöglicht. In dieser Version sind die meisten Daten verfügbar, jedoch mit einer Verzögerung von 15 Minuten. Das bedeutet, dass Nachrichtenartikel und Preise erst 15 Minuten nach ihrer Veröffentlichung über die API abrufbar sind. Echtzeitdaten stehen hingegen ausschließlich Premiumnutzern zur Verfügung.

Für das in dieser Arbeit entwickelte System stellen diese Einschränkungen jedoch kein Problem dar, da sämtliche benötigten Daten kostenlos zugänglich sind. Dazu zählen unter anderem historische Aktienpreise zu einem bestimmten Zeitpunkt mit Open-, High-, Low- und Close-Werten sowie technische Indikatoren wie Moving Averages oder der Relative Strength Index. Diese Datenquellen sind essenziell für die technische Analyse. Darüber hinaus bietet die API einen Endpunkt, der Nachrichten und die entsprechende Stimmung eines Unternehmens für einen ausgewählten Zeitraum bereitstellt, was für die Stimmungsanalyse genutzt wird. Die Entscheidung für Alphavantage basiert zudem auf der Tatsache, dass es sich um einen lizenzierten Partner der NASDAQ sowie einen Partner von etwa Amazon Web Services handelt. Dadurch wird eine hohe Datenqualität gewährleistet, ohne auf frei verfügbare, potenziell unzuverlässige Quellen im Internet zurückgreifen zu müssen.

3.1.2 ChatGPT API

Jede Anfrage an die API ist mit Kosten verbunden, die sich nach der Anzahl der genutzten Token bemessen. Beispielsweise belaufen sich die Kosten bei Nutzung des GPT-4-Modells auf 0,0025 USD pro

1.000 Token (entspricht etwa 750 englischen Wörtern) für den Input und 0,01 USD für den Output (Stand: 27.12.2024). Daher ist es von zentraler Bedeutung, die Anzahl der Token so effizient wie möglich zu gestalten, indem nur so viele wie nötig und so wenige wie möglich verwendet werden, um die Kosten gering zu halten.

In diesem Zusammenhang wird die „Text Generation“-API eingesetzt, welche es erlaubt, Anfragen mittels Prompts zu senden und entsprechende Ergebnisse in Textform zu erhalten. Darüber hinaus bietet die API die Möglichkeit, strukturierte Anfragen zu stellen und strukturierte Antworten im JSON-Format zu empfangen. Dies ist besonders für die automatisierte Nutzung von Vorteil, da die Antworten leicht durchlaufen und die enthaltenen Daten effizient weiterverarbeitet werden können.

Obwohl es Quellen gibt, die besagen, dass Finanzdaten akkurat von dem LLM wiedergegeben werden [18], werden in dieser Arbeit nicht die Daten von ChatGPT genutzt, sondern die von Alphavantage, da sich im folgenden Kapitel zeigen wird, dass dies nicht zielführend ist.

Im Rahmen der entwickelten Anwendung werden die von Alphavantage erhaltenen Daten an ChatGPT übermittelt, zusammen mit einem spezifischen Prompt, der die Verarbeitung der Daten gemäß dem jeweiligen Analyseansatz steuert. Die Antwort wird im JSON-Format zurückgegeben und anschließend in einer Datenbank gespeichert, um sie für weiterführende Analysen nutzen zu können.

3.1.3 Vergleich von ChatGPT und Alphavantage in Betracht der technischen Analyse

Um die Datenqualität der beiden in Betracht gezogenen Datenquellen zu vergleichen, wurden mehrere API-Endpunkte erstellt, die Daten entgegennehmen, wie zum Beispiel den Ticker der untersuchten Aktie, also das einzigartige Kürzel dieses Unternehmens an der Börse oder welches GPT-Modell genutzt werden soll (Abbildung 5).

The screenshot shows the Postman interface with the 'Run' tab selected. The URL is set to `http://localhost:3000/v1/comparison/news-sentiment-summary/{ticker}/{date}`. The 'Params' section shows six parameters: `ticker` (Value: AAPL, Type: string), `dateStart` (Value: 20231001T0001, Type: string), `dateEnd` (Value: 20231007T0000, Type: string), `model` (Value: gpt-4o, Type: string), `sort` (Value: RELEVANCE, Type: string). The 'Body' and 'Headers' sections are empty. The 'Auth' section is collapsed. The 'Pre Processors' and 'Post Processors' sections are also collapsed. The 'Settings' section is partially visible.

Abbildung 5: API-Referenz des Vergleichs von ChatGPT und Alphavantage

Es wurden die drei wichtigsten Merkmale verglichen: **Preisdaten**, **Daten für die technische Analyse** und das **Auswerten von Stimmungsanalysen** der von Alphavantage bereitgestellten Nachrichtenartikel. Beispieldaten des Endpunktes, die genutzten Prompts und Codeausschnitte sind in dem [GitHub-Repository](#) zu sehen.

In einer Untersuchung der Preisdaten stellte sich heraus, dass ChatGPT nicht in der Lage war, den exakten Kurswert korrekt wiederzugeben, obwohl die relevanten Daten innerhalb des Trainingszeitraums der untersuchten Modelle (GPT 4o und 4o Mini) lagen. Die Stichproben wichen weit von dem realen Wert von Alphavantage ab (Tabelle 1), besonders das Volumen war sehr weit von der Realität entfernt.

Date	Alphavantage				ChatGPT			
	Open	High	Low	Close	Open	High	Low	Close
2024-01-02	100.92	103.10	100.85	102.36	111.50	113.00	110.00	112.00
2024-01-03	102.27	103.62	101.66	103.22	112.00	114.00	111.00	113.50
2024-01-04	104.08	104.57	102.05	102.32	113.50	115.00	112.50	114.00
2024-01-05	103.17	103.39	102.13	102.63	-	-	-	-

Tabelle 1: Vergleich von Preisdatenanfrage

Um zu überprüfen, ob diese Diskrepanzen durch die Nutzung der ChatGPT-API und das vorgegebene JSON-Antwortschema verursacht wurden, wurden zusätzlich Anfragen direkt über die Plattform chatgpt.com unter Verwendung des Modells GPT 4o gestellt. In einer kleineren Stichprobe von 20 Anfragen waren die Ergebnisse deutlich genauer als über die API und oft richtig. Das liegt daran, dass der Chatbot im Internet nach Werten suchen kann und zum Beispiel Yahoo Finance dafür nutzt. Dieses Feature wird nicht in der API unterstützt.

Im Rahmen der Analyse technischer Daten wurde der Moving Average als Vergleichsgröße herangezogen. Dabei zeigte sich, dass ChatGPT-API nicht in der Lage war, diesen Wert korrekt zu berechnen. Auch über die Interaktion mit dem Chatbot auf *chatgpt.com* konnten keine validen Ergebnisse generiert werden. Es musste auf alternative Wege zurückgegriffen werden, da ChatGPT keine ungenauen Werte ausgeben wollte.

Erst durch die Formulierung des spezifischen Prompts „*If you had to tell me the values even though you are not sure, what would they be?*“, konnten geschätzte Werte vom Chatbot ermittelt werden. Diese erwiesen sich jedoch als ungenau und unzureichend für eine fundierte Analyse. Auch die Möglichkeit, ChatGPT zu veranlassen, Daten im Internet zu suchen, bot in diesem Fall keine hilfreiche Lösung. Dieses Ergebnis unterstützt die Aussagen, dass LLMs nicht gut darin sind, mathematische Aufgaben zu lösen [11].

Aufgrund dieser Einschränkungen und dem Fakt, dass teilweise bis zu 10 Sekunden länger auf eine Antwort von ChatGPT gewartet werden musste (Abbildung 15A), bestärkt sich das Argument eine externe, zuverlässige Quelle wie Alphavantage zu nutzen.

Abschließend wurde die Stimmungsanalyse der beiden Datenquellen bewertet. Da ChatGPT weder auf das Internet zugreifen kann noch in der Lage ist, eine Internetseite zu lesen, selbst wenn ein entsprechender Link zur Verfügung gestellt wird, wlich der sogenannte „SentimentScore“ – der von -1 (negativ) über 0 (neutral) bis 1 (positiv) reicht – erheblich von dem durch Alphavantage gelieferten SentimentScore ab. Ein möglicher Grund dafür könnte sein, dass die Lösung von Alphavantage den gesamten Artikel analysiert und nicht nur den Titel. Nach manueller Durchsicht von mehreren Nachrichtenfragen zum Beispiel von Google, Apple oder PayPal zeigte sich jedoch, dass auch der SentimentScore von Alphavantage nicht besonders präzise war. So wurde beispielsweise in einem Artikel vom Kauf der Apple-Aktie abgeraten, während Alphavantage einen leicht positiven SentimentScore von 0,2 berechnete. Dies legt die Vermutung nahe, dass auch Alphavantage lediglich den Titel und den ersten Satz, der als „Summary“ in der API bezeichnet wird, in die Analyse einbezieht. Auf die Nachfragen an Alphavantage, wie genau dieser SentimentScore ermittelt wird, gab es keine Antwort.

Obwohl es Studien gibt, die darauf hinweisen, dass allein die Analyse von Nachrichtentiteln aussagekräftige Ergebnisse liefern kann [56,57], liefert die Alphavantage-API gelegentlich Titel mit geringer Informationsdichte. Ein Beispiel dafür ist der Titel: „3 Things About Arm Holdings That Smart Investors Know“, der in einer Suche nach Apple-Nachrichten zurückgegeben wurde. Der zugehörige

Artikel beschreibt, dass „Arm Holdings“ ohne Apple wahrscheinlich nicht existieren würde, da Apple vor über 20 Jahren als einer der ersten Investoren eingestiegen ist und heute einer der größten Kunden ist. Zudem wird im Artikel erwähnt, dass der Marktanteil von ARM-Chips stetig wächst, was auch Apple zugutekäme [61]. Somit muss der gesamte Artikel gelesen werden, um für die Aktie relevante Informationen zu finden. Nur durch den Titel wäre unklar, was Apple mit diesem Nachrichtenartikel zu tun hat.

Für eine präzisere Berechnung des SentimentScores wäre es ideal, den gesamten Artikel zu analysieren. Dies ist jedoch kostenintensiv, da ein Artikel oft eine große Anzahl an Token umfasst, wodurch bei einer großen Anzahl an Artikeln die API-Kosten erheblich steigen, und es bestünde das Risiko, das Token-Limit pro Minute zu überschreiten, welches in dieser Untersuchung bei 30.000 Token für GPT-4o und 200.000 Token für GPT-4o-mini liegt, da Tier 1 genutzt wird [62]. Bei einer Artikellänge von 1000 Token wurden je nach Artikellänge teilweise nur drei Artikel von ChatGPT zurückgegeben und bewertet (Abbildung 22A). Mit 600 Token langen Artikeln können konstant mindestens acht Artikel analysiert werden.

```
{
  "title": "Jim Cramer's top 10 things to watch in the stock market Tuesday",
  "url": "https://www.cnbc.com/2024/10/01/jim-cramers-top-10-things-to-watch-in-the-stock-market-tuesday.html",
  "summary": "Wall Street is heading for a lower open on the first day of October.",
  "sentimentScore": "0.063848",
  "relevanceScore": "0.165409"
},
```

Abbildung 6: Stimmungsanalyse durch Alphavantage

```
{
  "title": "Jim Cramer's top 10 things to watch in the stock market Tuesday",
  "url": "https://www.cnbc.com/2024/10/01/jim-cramers-top-10-things-to-watch-in-the-stock-market-tuesday.html",
  "sentimentScore": "0.000000",
  "relevanceScore": "0.200000"
},
```

Abbildung 7: Stimmungsanalyse durch ChatGPT ohne Artikelinhalt

```
{
  "title": "Jim Cramer's top 10 things to watch in the stock market Tuesday",
  "url": "https://www.cnbc.com/2024/10/01/jim-cramers-top-10-things-to-watch-in-the-stock-market-tuesday.html",
  "summaryOfArticleContent": "Jim Cramer highlights that Boeing's stock is under pressure due to a potential $10 billion equity raise and ongoing worker strikes. The article notes that Wells Fargo has reduced its price target for Boeing, maintaining a sell-equivalent rating. These factors contribute to a negative outlook for Boeing in the near term. The focus on financial challenges and analyst downgrades results in a negative sentiment and high relevance to Boeing.",
  "sentimentScore": "-0.800000",
  "relevanceScore": "0.900000"
},
```

Abbildung 8: Stimmungsanalyse durch ChatGPT mit Artikelinhalt

Deng et al. [57] fanden in ihrer Studie heraus, dass die ersten Token eines Artikels die größte Aussagekraft besitzen, da durch den sogenannten „Lead Bias“ die wichtigsten Informationen meist am Anfang eines Artikels enthalten sind. Sie nutzten 100 Token, da jedoch durch die Verbesserung der Modelle die Nutzung günstiger wird, werden in diesem Programm die ersten 600 Token genutzt. Mit sieben analysierten Artikeln liegt der Preis bei ungefähr einem Cent. Durch das Bereitstellen der ersten 600 Token jedes Artikels, konnte die Genauigkeit des Relevanzwertes und des SentimentScores deutlich verbessert werden. Bei einem Artikel über Boeing haben Alphavantage und ChatGPT ohne Nachrichteninhalt einen relativ neutralen Relevanz- und SentimentScore (Abbildung 6 und 7), während ChatGPT mit Nachrichteninhalten und der Reasoning Prompting Strategie, einen hohen Relevanzscore hat, da Boeing direkt erwähnt wird und einen negativen SentimentScore hat, da ein schlechterer Aktienkurs erwartet wird. (Abbildung 8)

Zudem weisen die von Alphavantage bereitgestellten Nachrichtenartikel eine geringe Quellenvielfalt auf, da diese überwiegend von den Plattformen *benzinga.com* und *zacks.com* stammen. Diese eingeschränkte

Varianz mindert die inhaltliche Diversität der Informationen. Darüber hinaus enthalten die genannten Quellen gelegentlich Werbeanzeigen für mobile Applikationen, was die Seriosität und wissenschaftliche Verwertbarkeit dieser Inhalte infrage stellt. Ein weiterer Nachteil besteht darin, dass einige zurückgelieferte Artikel-URLs auf consent.google.com verweisen, wodurch die direkte Zugänglichkeit zu den entsprechenden Inhalten erheblich eingeschränkt wird. Aus diesen Gründen erfolgt die Beschaffung von Nachrichtenartikeln zusätzlich über die Plattform *GNews*, um eine höhere Quellenvielfalt und Verlässlichkeit sicherzustellen.

3.2 Technischer Aufbau des Handels- und Analysesystems

Um umfangreiche Datenbestände verarbeiten zu können, ist die Implementierung automatisierter Prozesse unerlässlich. Da zugleich nachträglich nachvollziehbar sein soll, welche Entscheidungen ChatGPT getroffen hat, ist die persistente Speicherung aller dem LLM zur Verfügung gestellten Informationen von zentraler Bedeutung. In diesem Kapitel wird daher zunächst die zugrunde liegende Datenbankstruktur erläutert, einschließlich der Gründe für deren spezifische Gestaltung. Anschließend folgt eine Darstellung der Softwarearchitektur, die den Fluss der Daten und Anfragen im Detail aufzeigt.

3.2.1 Datenbankstruktur

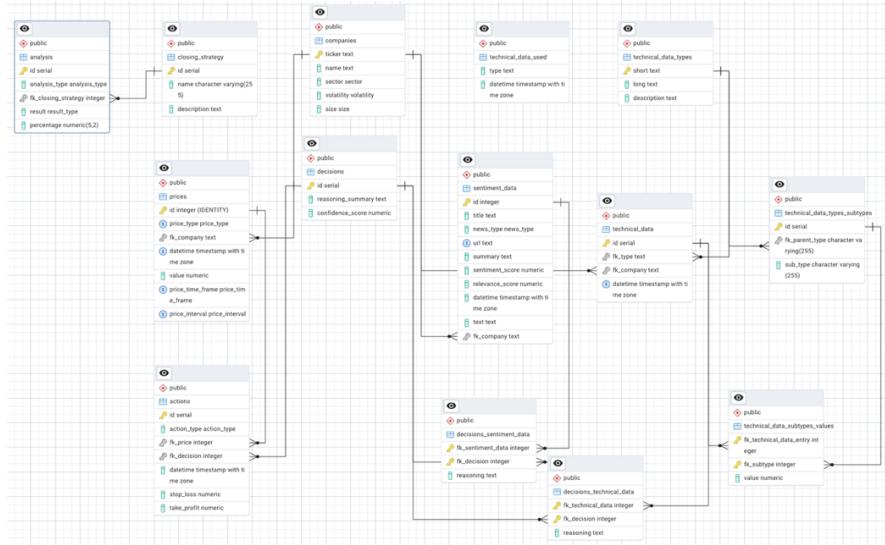


Abbildung 9: ERD der Datenbank

Die Tabelle **prices** speichert historische Aktienkurse rückblickend bis zurück in das Jahr 2000. Neben dem spezifischen Wert (*value*) des Kurses werden zusätzliche Attribute wie der Typ des Preises (*price_type*), der zeitliche Rahmen (*price_time_frame*), das Intervall (*price_interval*) sowie der Zeitstempel (*datetime*) erfasst. Diese Attribute erlauben es, die Datendichte flexibel zu handhaben, beispielsweise durch die Speicherung von stündlichen, täglichen oder wöchentlichen Kursen. Für die vorliegende Analyse werden ausschließlich tägliche Kursdaten betrachtet. Jede Preisangabe ist eindeutig definiert durch die Kombination aus *fk_company*, *price_time_frame*, *price_interval*, *datetime* und *price_type*.

Die Tabelle ***sentiment_data*** dient der Speicherung von Daten aus Stimmungsanalysen. Dort werden Nachrichtenartikel mit Informationen wie Titel (*title*), URL (*url*), einer Zusammenfassung (*summary*) und – falls verfügbar – dem gesamten Text (*text*) hinterlegt. Zusätzlich werden jedem Artikel ein Relevanz-Score (*relevance_score*) sowie ein Sentiment-Score (*sentiment_score*) zugewiesen. Diese Metriken bewerten die Bedeutung des Artikels für das jeweilige Unternehmen (*fk_company*) sowie die potenzielle Reaktion der Investoren auf dessen Inhalt. Die Kombination aus Unternehmenskennung und URL garantiert die Eindeutigkeit der Einträge.

Die technischen Daten werden in einer relationalen Struktur abgebildet, die der Vielgestaltigkeit technischer Indikatoren Rechnung trägt. In der Tabelle ***technical_data*** werden grundlegende technische Indikatoren mit ihrem Typ (*fk_type*), der zugehörigen Unternehmenskennung (*fk_company*) sowie einem Zeitstempel (*datetime*) gespeichert. Die Zuordnung zwischen Indikatoren und Subtypen erfolgt über die Tabelle ***technical_data_types_subtypes***, die beispielsweise Unterarten wie „Real Lower Band“, „Real Middle Band“ und „Real Upper Band“ für Bollinger-Bänder umfasst. Der tatsächliche Wert eines Subtyps wird schließlich in der Tabelle ***technical_data_subtypes_values*** gespeichert. Die Struktur stellt sicher, dass für jeden technischen Indikator alle zugehörigen Subtypen korrekt gruppiert und abrufbar sind, was insbesondere bei komplexeren Indikatoren wie Bollinger-Bändern erforderlich ist.

Die Tabelle ***actions*** dokumentiert die vom LLM getroffenen Investitionsentscheidungen. Diese umfassen die Art der Aktion (*action_type*), d. h. entweder eine Long-, Hold- oder Short-Position, sowie den Zeitpunkt der Entscheidung (*datetime*). Falls eine Aktion auf Basis technischer Analysen getroffen wurde, werden zusätzlich ein *stop_loss*-Wert und ein *take_profit*-Wert gespeichert.

Jede Aktion basiert auf Entscheidungen, die in der Tabelle ***decisions*** festgehalten werden. Hier wird eine zusammenfassende Begründung (*reasoning_summary*) sowie ein *confidence_score* gespeichert, der das Vertrauen des Modells in die Entscheidung quantifiziert. Die einzelnen Begründungen, die zu einer Entscheidung führten, werden in den Verknüpfungstabellen ***decisions_sentiment_data*** und ***decisions_technical_data*** gespeichert. Dabei wird für jeden genutzten Datenpunkt, sei es aus der Sentiment- oder technischen Analyse, ein separater Eintrag mit einer spezifischen Begründung erstellt. Diese Struktur soll das Modell zur detaillierten Argumentation anregen, da bekannt ist, dass dieser Ansatz die Performance von LLMs signifikant verbessern kann.

Die Tabelle ***closing_strategy*** ist eine Lookup-Tabelle, die die verschiedenen Strategien zum Schließen einer Position dokumentiert. Während die Tabelle ***actions*** lediglich die Eröffnung von Positionen abbildet, legt die ***closing_strategy*** fest, unter welchen Bedingungen diese geschlossen werden. Dies ermöglicht erst die Bewertung, ob ein Trade erfolgreich war oder nicht. Die Ergebnisse dieser Bewertungen werden in der Tabelle ***analysis*** gespeichert, die Informationen zur genutzten Schließstrategie (*fk_closing_strategy*), zum Ergebnis (*result*) sowie zur prozentualen Höhe des Gewinns oder Verlusts enthält.

Die Tabelle ***technical_data_used*** erfasst schließlich, wie häufig gewisse technische Indikatoren von den Algorithmen abgefragt wurden. Diese Informationen können dazu dienen, die Nutzungsmuster zu analysieren und die Effektivität verschiedener Indikatoren im Entscheidungsprozess zu bewerten.

3.2.2 Softwareaufbau und Datenfluss

Die Software basiert zentral auf einem Node.js-Backend, das mithilfe von Express.js API-Endpunkte bereitstellt, über die verschiedene Funktionalitäten aufgerufen werden können. Node.js wurde

ausgewählt, da JavaScript mit Python zum Zeitpunkt der Planung der Software die einzigen offiziell unterstützten Sprachen waren, die eine REST API Bibliothek von OpenAI zur Verfügung gestellt bekommen haben. Die Entscheidung ist auf JavaScript und somit Node.js gefallen, da das Ökosystem zum Erstellen einer API sehr einfach aufzubauen ist. Dabei muss aber erwähnt sein, dass dieses Backend in allen Sprachen gebaut werden hätte können.

Zu den Kernfunktionen zählen unter anderem die Verwaltung von in der Datenbank gespeicherten Unternehmen, welche analysiert werden, die Aggregation von Daten aus unterschiedlichen APIs sowie die Nutzung von ChatGPT. Direkte Prompt-Aufrufe aus eingehenden Anfragen sind nicht möglich. Stattdessen werden spezifische Funktionen ausgeführt, die vordefinierte Prompts an das LLM übermitteln.

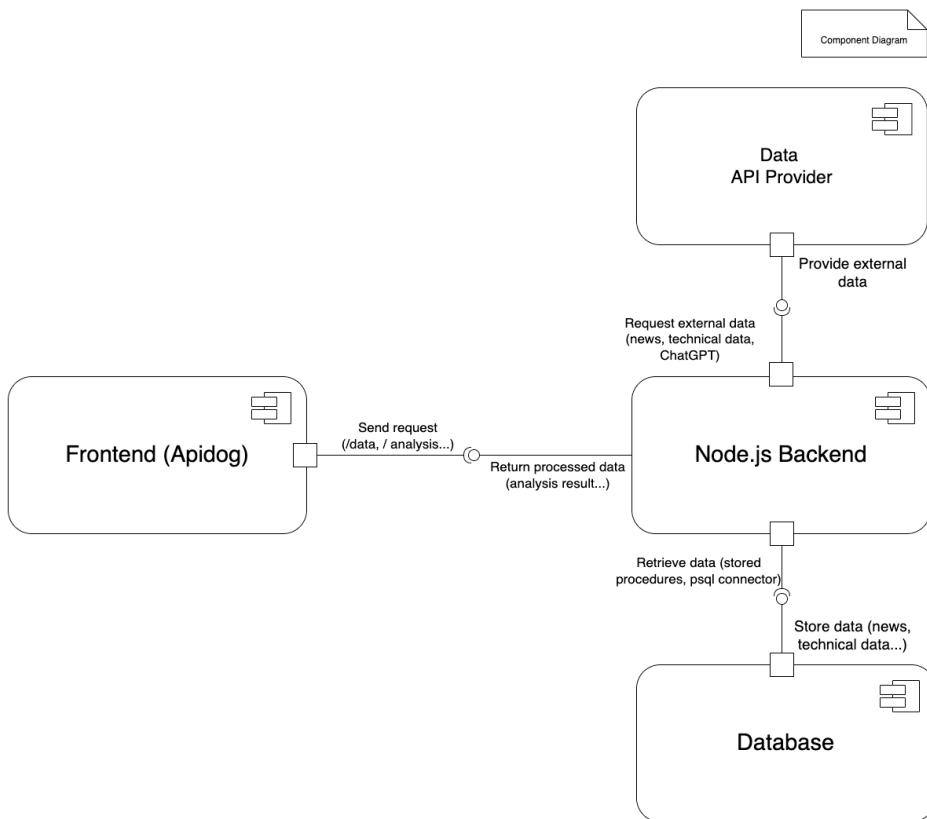


Abbildung 10: Softwareaufbau

Das Backend empfängt die API-Anfragen und delegiert diese an unterschiedliche Services, die entweder mit externen APIs interagieren, Daten in der Datenbank verwalten oder bestehende Datensätze abfragen. Zu diesen Services gehören beispielsweise `makeAlphaPriceRequest` oder `makeGPTToolRequest`, die durch die Übergabe spezifischer Parameter flexibel angepasst werden können. Dadurch ist es möglich, verschiedene Prompts innerhalb derselben Funktion zu verwenden, ohne redundanten Code für jeden Prompt zu implementieren.

Der Service `handleGPTResponse` ist so konzipiert, dass die Antworten von ChatGPT in die entsprechenden Tabellen der Datenbank übertragen werden. Dabei wird berücksichtigt, um welche Art der Analyse es sich handelt – sei es eine technische Analyse oder eine Stimmungsanalyse.

Die Prompts selbst sind nicht als feste Zeichenketten implementiert, sondern als Funktionen, die dynamisch zusammengesetzte Strings zurückgeben. Dies gewährleistet, dass identische Prompts für verschiedene Unternehmen und Zeiträume generiert werden können, wodurch eine Vergleichbarkeit der Antworten sichergestellt wird.

Unter dem Modul *types* befinden sich die Zod-Objekte (Abbildung 11), die an das LLM übergeben werden, um diesem vorzugeben, wie die Antworten im JSON-Format strukturiert sein sollen. Zod ist eine Schema-Validierungsbibliothek für JavaScript, die es ermöglicht, JSON-Schemas durch Funktionen zu definieren und beispielsweise Beschreibungen für einzelne Felder zu erstellen. Dadurch erhält das LLM eine präzisere Anleitung, wie die jeweiligen Felder zu handhaben sind.

Der entscheidende Vorteil dieser Methode liegt darin, dass das Schema nicht innerhalb des Prompts beschrieben werden muss, was häufig keine zuverlässige Einhaltung der gewünschten Datenstruktur gewährleistet. Stattdessen wird sichergestellt, dass Antworten stets im vorgesehenen Format zurückgegeben werden, wodurch die automatisierte Eintragung der Daten in die Datenbank ermöglicht wird. (Abbildung 12) Dennoch muss beachtet werden, dass trotz des definierten Schemas die Antworten nicht immer exakt dem erwarteten Format entsprechen. Daher ist es notwendig, vor der Übertragung in die Datenbank zu prüfen, ob alle erforderlichen Daten wie vorgesehen enthalten sind.

```
const generateReasoningResponse = () => {
  const reasoningDataSchema = z
    .object({
      id: z
        .number()
        .describe(
          "The id of the returned data from the functions that you use for the reasoning. F"
        ),
      reasoning: z
        .string()
        .describe(
          "One of the reasons that the model has for taking the action."
        ),
    })
    .describe("A reasoning behind the action");
  const reasonsArray = z
    .array(reasoningDataSchema)
    .describe(
      "An array of all the reasons behind the action. Many reasons can be given but should be s"
    );
  const responseFormat = z.object({
    reasons_array: reasonsArray,
    confidence_score: z
      .number()
      .refinement((data) => data >= 0 && data <= 1)
      .describe(
        "How confident is the model that the action will return a profit as a value between 0"
      ),
    reasoning_summary: z
      .string()
      .describe("A three sentence summary of the reasoning behind the action"),
    action: z
      .enum(["LONG", "SHORT", "HOLD"])
      .describe(
        "The action to take based on the reasoning. Long if buy, short if sell, hold if do no"
      ),
    stop_loss: z
      .number()
      .optional()
      .describe(
        "The price at which the stock should be sold if the price goes down. This is a percen"
      ),
    take_profit: z
      .number()
      .optional()
      .describe(
        "The price at which the trade should be closed. Only use this when using technical an"
      ),
  });
  return responseFormat;
};
```

Abbildung 11: Zod Antwortschema

```

    "result": {
      "reasons_array": [
        {
          "id": "20341",
          "reasoning": "The article discusses the positive outlook on home automation sys"
        },
        {
          "id": "20340",
          "reasoning": "This article highlights strong consumer engagement with a popular"
        },
        {
          "id": "20342",
          "reasoning": "The discussion of labor union challenges introduces potential neg"
        },
        {
          "id": "20343",
          "reasoning": "The article emphasizes the cultural significance of Prime Day, whi"
        },
        {
          "id": "20339",
          "reasoning": "Criticism of Amazon's Kindle for falling behind competitors introc"
        }
      ],
      "confidence_score": 0.7,
      "reasoning_summary": "The news articles present a mixed sentiment towards Amazon. While competitiveness. Overall, the positive sentiment outweighs the negative, suggesting a Lf"
      "action": "long"
    }
  }
}

```

Abbildung 12: Beispielantwort in gefordertem Format

Dieses Schema wird verwendet, wenn ChatGPT Vorschläge für Aktionen unterbreitet. Wie im Kapitel zum Datenbankaufbau beschrieben, soll das Feld *reasons_array* sicherstellen, dass ein reasoning-basierter Aspekt berücksichtigt wird, um die Qualität der Ergebnisse zu verbessern. Konkret bedeutet dies, dass ChatGPT für jedes Datum – sei es technischer oder stimmungsanalytischer Natur – ein Objekt erstellen soll, das aus der ID dieses Datums sowie einer Begründung besteht, warum dieses Datum genutzt wird und wie es zur Ergebnisfindung beiträgt.

Ein Problem dabei ist, dass ChatGPT gelegentlich fehlerhafte IDs erzeugt oder Begründungen liefert, die auf nichtexistente Daten beruhen. In einigen Fällen denkt sich das Modell sogar IDs aus, wenn zu einer Aktie in einem bestimmten Zeitraum keine Daten vorhanden sind, was insbesondere bei kleineren Unternehmen häufiger vorkommen kann. Nichtexistente IDs führen beim Einfügen in die Datenbank zu Fehlern, weshalb diese vorab gefiltert werden müssen.

Wie ersichtlich, sind Vorkehrungen erforderlich, um solche Fehler zu vermeiden. Allerdings bietet die JSON-Formatierung zumindest die Sicherheit, dass mit dem Objekt grundsätzlich gearbeitet werden kann und alle erforderlichen Felder vorhanden sind, auch wenn die Inhalte dieser Felder nicht immer korrekt sind. Ohne die festgelegte JSON-Struktur würde in Bezug auf diese beiden Aspekte eine Unsicherheit bestehen.

Es existiert außerdem ein Ordner namens *helper*, der verschiedene kleinere Methoden enthält, wie beispielsweise zur Validierung von Datumsformaten oder ähnlichen Aufgaben. Darüber hinaus umfasst dieser Ordner auch Tools, die für das sogenannte *Function Calling* genutzt werden. Dabei kann dem Modell neben dem Prompt und dem Antwortformat auch eine Menge an Tools übergeben werden. Dies erfolgt durch ein JSON-Objekt, das beschreibt, welche Methoden im Backend verfügbar sind und durch ChatGPT aufgerufen werden können (Abbildung 13).

Es ist jedoch wichtig zu betonen, dass das LLM die Methoden nicht direkt ausführt, da es keinen Zugriff auf das Backend hat. Stattdessen liefert es eine Antwort, bei der das Feld *'finish_reason'* mit *tool_calls* angegeben wird, anstatt mit *stop*. Dies ermöglicht eine Unterscheidung: Wenn der Grund für das Beenden der Antwort *stop* ist, handelt es sich um die finale Antwort. Wenn jedoch *tool_calls* als Grund angegeben ist, signalisiert dies, dass das LLM zusätzliche Informationen benötigt, um eine abschließende Aussage zu treffen.

```

const sentimentalTools = [
  {
    type: "function",
    function: {
      name: "get_news_data",
      description: "Get news articles with title, urls, summary, sentiment_score, relevance_score",
      parameters: {
        type: "object",
        properties: {
          ticker: {
            type: "string",
            description: "Ticker of the stock e.g. AAPL",
          },
          dateStart: {
            type: "string",
            description: "Start date of the asked news data in format 2024-04-10",
          },
          dateEnd: {
            type: "string",
            description: "End date of the asked news data in format 2024-04-19",
          },
        },
        required: ["ticker", "dateStart", "dateEnd"],
        additionalProperties: false,
      },
      strict: true,
    },
  },
];

```

Abbildung 13: Beschreibung der durch ChatGPT aufrufbaren Funktionen

Die Antwort mit der *finish_reason: tool_calls* enthält ein Array aus Objekten, die angeben, welche Methode mit welchen Parametern aufgerufen werden soll. Jedem Objekt wird dabei eine eindeutige ID zugeordnet. Im Backend kann dieses Array iterativ verarbeitet werden, wobei die entsprechenden Daten den Nachrichtenarrays mit der jeweiligen ID hinzugefügt werden. Dadurch wird sichergestellt, dass ChatGPT die Antworten den jeweiligen *tool_calls* korrekt zuordnen kann und die Informationen strukturiert verarbeitet werden.

3.3 Vorgehen bei der Simulation

ChatGPT führt keine Käufe oder Verkäufe direkt aus, sondern gibt lediglich Empfehlungen, welche Position – Long, Short oder Hold – an einem bestimmten Tag basierend auf den vorliegenden Daten eingegangen werden sollte. Dabei wird ChatGPT ein spezifisches Datum vorgegeben, und das Modell wird so geprompted, dass es "heute" dieses Datum ist und eine entsprechende Analyse für die angegebene Aktie durchgeführt. Damit das LLM keine Daten erhält, die den Preis oder weitere Informationen der Aktie preisgeben, wird bei den *tool calls* darauf geachtet, dass der Endzeitpunkt der abfragen maximal das Datum des zu analysierenden Tages ist.

Da der Zugriff auf externe APIs wie die Alphavantage API eingeschränkt ist, sofern keine kostenpflichtigen Pläne genutzt werden, erhält ChatGPT keine direkten API-Aufrufe für aktuelle Daten. Stattdessen werden nur vorab gesammelte Daten über mehrere Tage hinweg in die Analyse einbezogen. Zusätzlich muss auf die Kosten der ChatGPT-API-Aufrufe geachtet werden. Direkte API-Aufrufe durch ChatGPT könnten zu umfangreichen Antworten führen, wodurch der Preis pro Anfrage durch eine erhöhte Token Anzahl deutlich steigen würde. Die Einschränkung auf vorgegebene Services und Zeiträume ermöglicht eine gezielte Steuerung der Token Anzahl und damit der Kosten.

Darüber hinaus bietet das Speichern der verwendeten Daten in der Datenbank den Vorteil, dass nachvollziehbar bleibt, welche Informationen konkret in die Entscheidungen eingeflossen sind und wie diese zu den Empfehlungen beigetragen haben. Konkret bedeutet dies, dass alle Daten, die ChatGPT für die Analyse zur Verfügung stehen sollen, zunächst in der Datenbank abgelegt werden müssen, bevor ein

umfassender Analyselauf gestartet wird. In diesem Lauf werden alle zu untersuchenden Aktien für einen vorgegebenen Zeitraum analysiert.

Bei der Stimmungsanalyse erfolgt der Prozess in zwei Schritten. Zunächst werden die gecrawlten Nachrichtenartikel mit einer spezifischen Prompt an ChatGPT übergeben, um für jeden Artikel einen relevance_score und einen sentiment_score zu erzeugen. Diese Scores werden in der Datenbank gespeichert. Im zweiten Schritt – während der eigentlichen Analyse – werden diese Werte über die tool calls an ChatGPT übermittelt. ChatGPT greift in diesem Schritt nicht mehr direkt auf die Artikeltexte zu, sondern verwendet ausschließlich die zuvor erstellten Zusammenfassungen und Scores. Diese Vorgehensweise hat den Vorteil, dass die Scores auf Basis der gesamten Artikelinhalte erstellt werden, was zwar initial zeit- und ressourcenintensiv ist, die eigentliche Analyse jedoch erheblich effizienter macht. Die reduzierten Datenmengen gewährleisten, dass das Wissen des gesamten Artikels in die Scores integriert ist, ohne dass ChatGPT die vollständigen Texte erneut verarbeiten muss. Die technische Analyse hingegen umfasst nur einen Schritt, bei dem ChatGPT während der Analyse alle Daten zu Verfügung gestellt bekommt.

In den folgenden Kapiteln wird erklärt, wie die Strategien einen Long oder Short Trade eingehen.

3.3.1 Technische Analyse

Das LLM verfügt über die Möglichkeit, mithilfe des Tool Callings eigenständig jene Daten abzurufen, die für seine Analyse relevant erscheinen. Dadurch kann das Modell je nach Marktbedingungen und vorgefundenen Preisstrukturen zwischen unterschiedlichen Verfahren der technischen Analyse wählen, beispielsweise zwischen einem gleitenden Durchschnitt (Moving Average), dem RSI (Relative Strength Index) oder dem Average Directional Movement Index (ADX). Diese Flexibilität ist bewusst vorgesehen, um es dem LLM zu ermöglichen, entsprechend der spezifischen Charakteristika einer Aktie und ihrer Preisentwicklung die am besten geeignete Analysemethode zu wählen.

Der folgende Prompt wird intern genutzt, um einen Handelsempfehlungsprozess durchzuführen:

„The current date (todays date) is \${currentDate}. You are supposed to tell if the stock \${symbol} is a LONG or SHORT for the current date based on the technical data that you can use by function calling. You can access technical data for the last 30 days and price data for the last 90 days.

Use the most fitting technical analysis strategies for the current price trend e.g. Moving Averages Crossover, Bollinger Bands narrowing, Support, Resistance and more. Reason what you see in the data and suggest a LONG position if you see a positive trend and a SHORT position if you see a negative trend. Else HOLD.

Also specify what the stop_loss percentage should be (value between 0 and 1) and what the exact take_profit price is with 4 decimal places. Explain why you chose these values based on the technical and price data by using Chain of Thought.“

Durch diese Formulierung sind bereits einige gängige Analysemethoden beispielhaft aufgeführt, sodass das Modell sich daran orientieren kann. Zudem wird das Modell explizit dazu angehalten, die Entscheidungsfindung offenzulegen und eine erläuternde Begründung anzugeben, da es als Best-Practice-Ansatz im Prompt Engineering gilt.

Nach der Analyse sämtlicher angeforderter Daten generiert das Modell eine Handlungsempfehlung in Form einer Long- oder Short-Position. Darüber hinaus gibt es eine Stop-Loss-Marke an (als prozentualer Wert in Bezug auf den aktuellen Kurs) sowie ein absolutes Take-Profit-Kursziel. Die Stop-Loss-Marke beschreibt, wie weit sich der Kurs entgegen der erwarteten Richtung entwickeln darf, bevor die Position zur Verlustminimierung automatisch geschlossen wird. So würde bei einem aktuellen Aktienkurs von 100 US-Dollar und einem Long-Trade mit Stop-Loss von 0,02 (2 %) die Position bei einem Kurs von 98 US-Dollar geschlossen werden, was einen Verlust von 2 % sind. Umgekehrt bezeichnet das Take-Profit-Limit jenen absoluten Kurs, bei dessen Erreichen die Position automatisch geschlossen wird, sobald die erwartete Kursbewegung eintritt. Beispielsweise würde bei einem Short-Trade bei 100 US-Dollar und einem definierten Take-Profit bei 95 US-Dollar die Position bei Erreichen von 95 US-Dollar realisiert. Sollte der Kurs hingegen auf 102 US-Dollar steigen und ein Stop-Loss von 0,02 aktiviert sein, würde die Position bei einem Verlust von 2 % geschlossen.

3.3.2 Stimmungsanalyse

In der Stimmungsanalyse ist es nicht üblich, Stop-Loss- oder Take-Profit-Werte anzugeben. Der Grund hierfür liegt darin, dass aus der Stimmungsanalyse allein keine Rückschlüsse auf die Intensität oder potenzielle Umkehrpunkte des Trends gezogen werden können. Folglich konzentriert sich diese Methodik ausschließlich auf die Frage, ob eine Long-, Short- oder Hold-Position in Erwägung zu ziehen wäre. Die interne Prompt-Struktur gestaltet sich daher wie folgt:

“The current date (todays date) is \${currentDate}. You are supposed to tell if the stock \${symbol} is a LONG or SHORT for the current date based on the news articles of today and yesterday.

*When the overall sentiment of the news articles is positive enough, you should suggest a LONG position.
When the overall sentiment of the news articles is negative enough, you should suggest a SHORT position. Else HOLD.”*

In der untersuchten Literatur zu diesem Themengebiet wurden in vielen Fällen einzelne Nachrichtenartikel herangezogen, um auf Basis der jeweiligen Stimmungsbewertungen eine Long- oder Short-Position zu eröffnen [55]. Andere Ansätze sammelten mehrere Artikel und leiteten eine Handelsentscheidung daraus ab, indem nach einer Aggregation der Stimmungsbewertungen bei überwiegend positiven Inhalten eine Long- und bei überwiegend negativen Inhalten eine Short-Position eingegangen wurde [57].

Im Rahmen der vorliegenden Untersuchung wird jedoch eine andere Methodik angewandt, die das erweiterte Kontextverständnis und das große Kontextfenster von ChatGPT nutzt. Hierbei ruft das Modell mithilfe des Tool Callings die relevanten Nachrichtenartikel des aktuellen und des vorangegangenen Tages ab. Diese Artikel liegen bereits mit einem Sentiment-Score vor, wobei nur die für den Untersuchungszeitraum relevantesten Artikel herangezogen werden. Durch die gemeinsame Auswertung aller Artikel ist das Modell nicht auf eine einzelne Quelle beschränkt, sondern kann sämtliche Beiträge im Kontext zueinander gewichtet und bewertet. In dem ersten Schritt der Stimmungsanalyse werden die Artikel jeweils einzeln untersucht, was schließlich die Grundlage für die übergreifende Bewertung bildet, um eine stärkere Empfehlung für eine Long- oder Short-Position abzuleiten.

3.3.3 Realisierung von Gewinnen und Verlusten

Wie bereits in den vorangegangenen Kapiteln dargelegt, generiert ChatGPT ausschließlich Handlungsempfehlungen (Long, Short oder Hold), tätigt jedoch keine tatsächlichen Kauf- oder Verkaufsorders. Um dennoch die Güte der von ChatGPT ausgesprochenen Empfehlungen zu beurteilen, kommt ein nachträgliches Evaluationsverfahren zum Einsatz.

Da die Datenbank Preisverläufe ab dem Jahr 2000 bis Anfang 2025 enthält und bekannt ist, an welchem Datum ChatGPT die jeweilige Empfehlung ausgesprochen hat, lässt sich die Preisentwicklung ab dem entsprechenden Eröffnungstag untersuchen. Dieses Verfahren wird mittels verschiedener Trading-Strategien realisiert. Für jeden Datensatz in der *actions*-Tabelle finden sich zugehörige Strategien in der Tabelle *closing_strategy*. Bei technischen Analysen wird der Kursverlauf seit Öffnung des Trades herangezogen. Das Schließen des Trades erfolgt dabei, wenn ein Stop-Loss oder ein Take-Profit ausgelöst wird.

Ein Problem besteht darin, dass in der *prices*-Tabelle lediglich tägliche Open-, High-, Low- und Close-Kurse gespeichert sind. Wenn beispielsweise ein Take-Profit bei 135 US-Dollar erreicht wird, der Tages-Schlusskurs jedoch bei 137 US-Dollar liegt, kann die ausschließliche Verwendung des Schlusskurses zu einer verzerrten Renditeberechnung führen. Um dieses Problem zu vermeiden, wird stattdessen der exakte Ausstiegskurs verwendet, der dem festgelegten Stop-Loss oder Take-Profit entspricht. Konkret bedeutet dies, dass zur Bestimmung des realisierten Ausstiegskurses entweder der absolute Dollarwert des Take-Profits oder der Kaufkurs abzüglich des vordefinierten Stop-Loss-Prozentsatzes herangezogen wird. In Fällen, in denen weder der Take-Profit erreicht wurde noch der Stop-Loss ausgelöst hat, bleibt der Trade aktiv und eine abschließende Beurteilung ist noch nicht möglich.

Für andere Strategien, die nicht auf Stop-Loss oder Take-Profit basieren, wie beispielsweise *days_n_close* (wobei $n \leq 30$), wird der Trade am Ende des n -ten Handelstages nach der Eröffnung automatisch geschlossen. Hiermit lassen sich ebenfalls Aussagen über die Qualität der Empfehlung treffen. Da die Handelsempfehlungen von ChatGPT bei dieser Strategie immer innerhalb des definierten Zeitraums enden, lassen sich rückblickend alle Trades abschließend bewerten.

In Abbildung 14 ist das PostgreSQL Prozedur dargestellt, welches für das Auswerten der Analysen ausgeführt wird. In dieser Prozedur werden zunächst in einer äußeren Schleife alle Einträge der Tabelle *actions* durchlaufen. Jeder Datensatz enthält Informationen zum Eröffnungskurs (Open-Preis) sowie das zugehörige Aktiensymbol (Ticker). Anschließend iteriert eine innere Schleife über alle hinterlegten Schließungsstrategien und unterscheidet dabei zwei Szenarien:

Liegt die ID der jeweiligen Strategie bei 6 (in diesem Fall die Take-Profit-Strategie), wird eine Subprozedur aufgerufen, die anhand der in *actions* definierten Take-Profit- und Stop-Loss-Grenzwerte den Schlusskurs (Closing-Preis) ermittelt.

Sollte die ID hingegen nicht 6 sein, wird aus dem Strategienamen der Wert n ausgelesen, der als Parameter an eine weitere Subprozedur übergeben wird. Diese ermittelt den Closing-Preis n Tage nach Eröffnung des jeweiligen Trades.

Basierend auf dem so berechneten Kursverlauf wird die Rendite je nach Positionstyp unterschiedlich bestimmt. Bei einer Long-Position wird die Differenz aus Closing- und Open-Preis berechnet, was bei steigenden Kursen zu einer positiven Rendite führt und bei fallenden Kursen zu einer negativen Rendite. Im Falle einer Short-Position wird der Open-Preis vom Closing-Preis subtrahiert, um den Gewinn oder

Verlust korrekt abzubilden. Abschließend speichert die Prozedur sämtliche Ergebnisse in der Tabelle *analysis*.

```

CREATE OR REPLACE PROCEDURE public.apply_closing_strategies()
LANGUAGE plpgsql
AS $procedure$
DECLARE
    act    RECORD;
    strat  RECORD;
    v_close_price  NUMERIC;
    v_percentage   NUMERIC(5,2);
    v_result       result_type;
    v_days_to_close INT;
BEGIN
    FOR act IN
        SELECT a.*,
               p.value      AS open_price_value,
               p.fk_company AS ticker
        FROM actions a
        JOIN prices p ON p.id = a.fk_price
    LOOP
        IF act.action_type = 'hold' THEN
            CONTINUE;
        END IF;

        FOR strat IN
            SELECT * FROM closing_strategy ORDER BY id
        LOOP
            IF strat.id = 6 THEN
                IF act.stop_loss IS NULL OR act.take_profit IS NULL THEN
                    CONTINUE;
                END IF;

                v_close_price := get_stop_take_exit_price(
                    p_ticker      => act.ticker,
                    p_open_date   => act.datetime,
                    p_long_or_short => act.action_type,
                    p_stop_loss   => act.stop_loss,
                    p_take_profit => act.take_profit,
                    p_open_price  => act.open_price_value,
                    p_price_type  => 'close',
                    p_time_frame  => 'TIME_SERIES_DAILY'
                );

                IF v_close_price IS NULL THEN
                    CONTINUE;
                END IF;

            ELSE
                IF strat.name LIKE 'day_%_close' THEN
                    v_days_to_close := CAST(
                        | split_part(split_part(strat.name, '_', 2), '_', 1) AS INT
                    );
                ELSE
                    v_days_to_close := 0;
                END IF;

                v_close_price := get_nth_trading_day_price(
                    p_ticker      => act.ticker,
                    p_open_date   => act.datetime,
                    p_n          => v_days_to_close,
                    p_price_type  => 'close',
                    p_time_frame  => 'TIME_SERIES_DAILY'
                );

                IF v_close_price IS NULL THEN
                    CONTINUE;
                END IF;
            END IF;

            IF act.action_type = 'long' THEN
                v_percentage := ROUND(
                    ((v_close_price - act.open_price_value)
                     / act.open_price_value) * 100,
                    2
                );
            ELSIF act.action_type = 'short' THEN
                v_percentage := ROUND(
                    ((act.open_price_value - v_close_price)
                     / act.open_price_value) * 100,
                    2
                );
            END IF;
        END IF;

        v_result := CASE WHEN v_percentage >= 0 THEN 'WIN' ELSE 'LOSS' END;

        INSERT INTO analysis(
            fk_action,
            analysis_type,
            fk_closing_strategy,
            result,
            percentage
        )
        VALUES(
            act.id,
            act.analysis_type,
            strat.id,
            v_result,
            v_percentage
        );
    END LOOP;
END;
$procedure$
```

Abbildung 14: Prozedur zum Auswerten der Handelsempfehlungen

3.4 Untersuchte Unternehmen

In der vorliegenden Arbeit wird der Fokus ausschließlich auf US-amerikanische Aktien gelegt. Dies erfolgt vor dem Hintergrund, dass die Mehrheit, der in der Literatur identifizierten Untersuchungen ebenfalls US-amerikanische Aktien analysiert. Um eine Verzerrung der Ergebnisse durch eine einseitige Betrachtung, beispielsweise ausschließlich von Technologieaktien, zu vermeiden, wurden drei Kriterien definiert, anhand derer eine ausgewogene Auswahl von Unternehmen getroffen wurde.

Erstens wurde der Sektor, in dem die Unternehmen tätig sind, berücksichtigt. Dabei wurden die Sektoren Technologie, Konsumgüter und Dienstleistungen, Energie und Industrie, Gesundheitswesen sowie Finanzwesen einbezogen. Zweitens wurde die Volatilität der Aktien herangezogen, welche in die Kategorien niedrig, mittel und hoch unterteilt ist. Die Einordnung erfolgt auf Basis der annualisierten Standardabweichung: Werte unter 20 % werden als niedrige Volatilität, Werte zwischen 20 % und 40 % als mittlere Volatilität und Werte über 40 % als hohe Volatilität klassifiziert. Drittens wurde die Unternehmensgröße bewertet, gemessen anhand der Marktkapitalisierung. Unternehmen mit einer Marktkapitalisierung von bis zu 2 Milliarden US-Dollar gelten als klein, solche mit einer Marktkapitalisierung zwischen 2 und 10 Milliarden US-Dollar als mittelgroß, und Unternehmen mit einer Marktkapitalisierung über 10 Milliarden US-Dollar werden als groß eingestuft. Somit kann auch der Einfluss von potenziell unterschiedlicher Nachrichtendichte untersucht werden. Die Standardabweichung wurde auf Grundlage von Daten der letzten drei Jahre der jeweiligen Unternehmen ermittelt, die über die Alphavantage-API abgerufen wurden. Informationen zu den Sektoren und der Marktkapitalisierung stammen von Yahoo Finance und entsprechen dem Stand vom 8. Januar 2025. Theoretisch wäre es möglich, sämtliche Unternehmen zu analysieren, da die Software den gesamten Prozess automatisieren kann. In der Praxis sind die Möglichkeiten jedoch begrenzt, da für die Datenerhebung die kostenlose API von Alphavantage verwendet wird, welche eine Begrenzung in der Anzahl der Abfragen aufweist. Zudem sollen die Abfragen der ChatGPT-API möglichst effizient gestaltet werden, um höhere Kosten zu vermeiden. Deswegen werden nur die Unternehmen aus Tabelle 1 untersucht.

Ticker	Name	Sektor	Volatilität	Größe
AAPL	Apple	technology	medium	large
AMPL	Amplitude	technology	high	small
AMZN	Amazon	consumer goods and services	high	large
AOS	A. O. Smith	energy and industry	medium	medium
CDMO	Avid Bioservices	healthcare	high	small
CSCO	Cisco	technology	medium	large
DIS	Disney	consumer goods and services	medium	large
DKNG	DraftKings	consumer goods and services	high	medium
EBF	Ennis	consumer goods and services	medium	small
FSLR	First Solar	energy and industry	high	large
GOEV	Canoo	energy and industry	high	small
IBIO	iBio	healthcare	high	small
JNJ	Johnson & Johnson	healthcare	low	large
META	Meta Platforms	technology	high	large
MGM	MGM	consumer goods and services	medium	medium
MRK	Merck & Co.	healthcare	low	large
NHI	National Health Investors	healthcare	medium	medium
NVDA	NVIDIA	technology	high	large
ORI	Old Republic International	consumer goods and services	medium	medium
OUST	Ouster	technology	high	small
PG	Procter & Gamble	consumer goods and services	low	large
PL	Planet Labs PCB	technology	high	small
QGEN	Qiagen	technology	medium	medium
ROKU	Roku	consumer goods and services	high	medium
T	AT&T	technology	medium	large
TSLA	Tesla	technology	high	large
VSTM	Verastem	healthcare	high	small
W	Wayfair	consumer goods and services	high	medium
WMT	Walmart	consumer goods and services	medium	large
XOM	Exxon	energy and industry	medium	large

Tabelle 2: Untersuchte Unternehmen nach Sektor, Volatilität und Marktkapitalisierung

3.5 Bewertungskriterien und Zielparameter

Für die Beurteilung der in der Einleitung aufgeworfenen Fragestellung, inwiefern ChatGPT durch Stimmungsanalyse oder technische Analyse geeignetere Aktienprognosen liefert, werden zunächst zentrale Bewertungskriterien definiert. Dabei steht im Vordergrund, wie genau die Modellvorhersagen sind (**Genauigkeit**), in welchem Umfang sie reproduzierbar sind (**Konsistenz**) und wie viel Zeit oder Ressourcen (**Effizienz**) für die Erzeugung einer verlässlichen Prognose benötigt werden. Ein hoher Grad an Genauigkeit ist unerlässlich, da sich Abweichungen zwischen erwartetem und realem Aktienkurs in der Rendite widerspiegeln. Konsistenz spielt eine ebenso wichtige Rolle, um sicherzustellen, dass bei wiederholter Anwendung desselben Prompt- und Dateninputs ähnliche Empfehlungen generiert werden. Außerdem fließt die Effizienz in die Bewertung ein, denn eine Lösung, die zwar gute Ergebnisse liefert, aber unverhältnismäßig hohe Kosten verursacht, wäre für den praktischen Einsatz nicht interessant. Neben diesen Faktoren ist auch die **Praktikabilität** entscheidend, also die Frage, ob und wie leicht die gewählten Methoden in eine reale Handlung überführt werden könnte. Um die Ergebnisse der stimmungsbasierten und der technischen Daten beruhenden Prognosen systematisch miteinander vergleichen zu können, müssen die Bewertungen mithilfe klarerer und einheitlicher Indikatoren erfolgen.

Für die **Genauigkeit** werden die *Gewinn-Verlust-Quote* und die *durchschnittliche Rendite pro Trade* untersucht. Die Gewinn-Verlust-Quote liefert Aufschluss darüber, wie viele Trades erfolgreich im Vergleich zu verlustreichen Trades verlaufen. Damit lässt sich erkennen, ob das Analyseverfahren tendenziell gewinnbringende oder eher verlustbehaftete Entscheidungen trifft. Die durchschnittliche Rendite pro Trade ergänzt diese Betrachtung, indem sie Auskunft über das tatsächliche Ausmaß der Gewinne oder Verluste gibt. Darüber hinaus wird die *Gesamtrendite für das Jahr 2024* betrachtet, um die kumulative Entwicklung in dem definierten Zeitraum zu erfassen und ein langfristiges Bild über die Profitabilität der Verfahren zu erhalten und auch in anderen Untersuchungen genutzt wird [63]. Ein weiterer Indikator für die Genauigkeit ist die Trefferquote, also welcher Anteil der Trades gewonnen wurde. Diese Metrik wird in vielen Untersuchungen genutzt, um die Genauigkeit eines Modells zu bewerten [15,55].

Die **Konsistenz** wird anhand der *Reproduzierbarkeit der Ergebnisse* und der *Standardabweichung der Handelsergebnisse* untersucht. Eine hohe Reproduzierbarkeit setzt voraus, dass bei identischen Eingangsdaten möglichst ähnliche oder sogar identische Trades generiert werden. Die Standardabweichung der Handelsergebnisse bezieht sich auf die Streuung der Rendite, gemessen über mehrere Durchläufe unter den gleichen Bedingungen. Geringe Schwankungen deuten dabei auf eine robuste und gleichbleibende Leistungsfähigkeit der Vorhersage hin. Konsistenz ist im Kontext des Tradings von besonderer Bedeutung, da selbst kleine Abweichungen bei den gewinnenden Trades einen negativen Einfluss auf die Gesamtrendite haben können [64].

Zur Bewertung der **Effizienz** werden die *Rechenzeit* und die *Kosten pro Prognose* herangezogen. Eine kurze Rechenzeit ist insbesondere dann essenziell, wenn die Modelle in Echtzeit oder zumindest zeitnah zum Marktgeschehen agieren sollen, während die Kosten pro Prognose eine wichtige Rolle spielen, wenn die Lösung wirtschaftlich nachhaltig betrieben werden soll.

Abschließend wird die **Praktikabilität** anhand der *Handhabung für den Anwender* und der *Integrationsfähigkeit in bestehende Systeme* bewertet. Eine unkomplizierte Bedienung trägt erheblich dazu bei, dass die Methode auch von Nutzern mit wenig technischem Hintergrundwissen erfolgreich eingesetzt werden kann. Zudem ist die Möglichkeit einer nahtlosen Anbindung an vorhandene

Infrastruktur, etwa durch standardisierte Schnittstellen, wichtig, um den produktiven Einsatz des Analyseverfahrens sicherzustellen.

4 Analyse der Daten

In diesem Kapitel werden die gesammelten Daten anhand der in den vorangegangenen Kapiteln definierten Bewertungskriterien und Zielparameter untersucht und miteinander verglichen. Zunächst wird die Konsistenz analysiert, gefolgt von der Untersuchung der Effizienz. Anschließend wird die Praktikabilität bewertet, bevor abschließend die Genauigkeit der Ergebnisse geprüft wird.

4.1 Konsistenz

Dieses Kapitel befasst sich mit der Reproduzierbarkeit der Analysen anhand der Standardabweichung der Handelsergebnisse, die als zentrale Kriterien für die Bewertung der Konsistenz herangezogen werden. Es stellt sich heraus, dass keine der beiden Analyseverfahren konsistent genug sind, um über mehrere Wiederholungen die gleichen Ergebnisse zu erhalten.

Zur Untersuchung der Reproduzierbarkeit wurden fünf Unternehmen ausgewählt, für die innerhalb einer gesamten Handelswoche Empfehlungen von ChatGPT sowohl basieren auf der Stimmungs- als auch auf der technischen Analyse erstellt wurden. Dieser Vorgang wurde insgesamt fünfmal wiederholt, um etwaige Abweichungen in den generierten Empfehlungen zu erfassen. Anschließend wurde erhoben, wie häufig in jedem Durchlauf Long-, Short oder Hold-Positionen empfohlen wurden, um die Konsistenz der Handelsempfehlungen zu bewerten (Tabelle 2).

In den fünf Durchläufen lassen sich nur geringe Abweichungen in den generierten Empfehlungen erkennen. Insbesondere bei der Stimmungsanalyse, welche die Empfehlung von Hold-Positionen vorsieht, ist zu beobachten, dass bei unzureichenden Informationen konsequent ein Hold vorgeschlagen wird. Dieser Befund ist insofern positiv zu werten, als Nachrichtenartikel überwiegend zu großen Technologieunternehmen wie AAPL oder TSLA verfügbar sind, während es für anderen Unternehmen und Zeiträume deutlich weniger relevante Meldungen gibt.

Auch für die technische Analyse zeigt sich eine ähnlich hohe Konstanz. Insgesamt variiert die Anzahl der Empfehlungen für die fünf betrachteten Unternehmen in den einzelnen Durchläufen um höchstens eins. Dies weist darauf hin, dass ChatGPT die vorhandenen Daten konsistent abfragt und einheitlich interpretiert.

Auf den ersten Blick erscheint die geringe Abweichung der Ergebnisse als positiv. Insbesondere zeigt sich in Abbildung 17A bis Abbildung 21A, dass die Standardabweichung der gewonnenen und verlorenen Trades beim Vergleich innerhalb der einzelnen Strategien maximal 1,26 beträgt. Allerdings wird bei Betrachtung der gesamten Anzahl an gewonnenen und verlorenen Trades deutlich, dass die Gesamtstreuung der Ergebnisse größer ausfällt. Bereits bei der vorliegenden, kleinen Stichprobe beträgt die maximal Standardabweichung 3,01. Es ist daher anzunehmen, dass sich diese Abweichung bei einer größeren Datenbasis weiter erhöhen würde.

Daher lässt sich feststellen, dass die Reproduzierbarkeit des Experiments grundsätzlich als einfach einzustufen ist, da die Prompts in den Funktionen stets identisch aufgerufen werden. Allerdings zeigt sich insbesondere bei kleineren Unternehmen eine mangelnde Konsistenz der generierten Handelsempfehlungen. So können bei unterschiedlichen Zeitpunkten der Abfrage trotz identischer

Datenbasis abweichende Ergebnisse entstehen. Die einzige durchgehend konsistente Beobachtung ist, dass ChatGPT bei unzureichender Datenlage keine Handelsempfehlung ausspricht.

Unternehmen	analysis_type	action_type	run 1	run 2	run 3	run 4	run 5
AAPL	technical	long	5	5	5	5	5
	technical	short	0	0	0	0	0
	technical	hold	0	0	0	0	0
AAPL	sentiment	long	4	3	3	4	2
	sentiment	short	0	2	1	0	1
	sentiment	hold	1	0	1	1	2
DKNG	technical	long	4	4	4	3	4
	technical	short	1	1	1	2	1
	technical	hold	0	0	0	0	0
DKNG	sentiment	long	1	1	1	1	1
	sentiment	short	0	0	0	0	0
	sentiment	hold	4	4	4	4	4
IBIO	technical	long	1	0	1	0	1
	technical	short	4	5	4	5	4
	technical	hold	0	0	0	0	0
IBIO	sentiment	long	0	0	0	0	0
	sentiment	short	0	0	0	0	0
	sentiment	hold	5	5	5	5	5
MRK	technical	long	2	2	2	2	1
	technical	short	3	3	2	3	3
	technical	hold	0	0	1	0	1
MRK	sentiment	long	0	0	0	0	0
	sentiment	short	0	0	0	0	0
	sentiment	hold	5	5	5	5	5
XOM	technical	long	5	5	5	5	5
	technical	short	0	0	0	0	0
	technical	hold	0	0	0	0	0
XOM	sentiment	long	0	0	0	0	0
	sentiment	short	0	0	0	0	0
	sentiment	hold	5	5	5	5	5

Tabelle 3: Anzahl der Handelsempfehlungen pro Unternehmen, Analysetyp und Handelsempfehlung

4.2 Effizienz

Die Effizienz der verschiedenen Analysemethoden wird anhand der Rechenzeit sowie der Kosten pro Prognose verglichen und bewertet. Dabei wird bei jedem Aufruf der ChatGPT-API die benötigte Zeit gemessen. Die Messung beginnt unmittelbar vor dem Absenden des Prompts und endet nach der letzten Antwort von ChatGPT. Somit wird nicht die Dauer einzelner Antwortvorgänge erfasst, sondern die gesamte Zeit, die ChatGPT für die Analyse benötigt, einschließlich etwaiger Tool-Aufrufe zur Datenbeschaffung. Auch Datenbankabfragen sind in dieser Zeitspanne enthalten, wenngleich ihr Einfluss mit einer durchschnittlichen Dauer von etwa 25 Millisekunden auf einem MacBook Air M1 als

vernachlässigbar betrachtet werden kann. Es wird deutlich, dass obwohl die technische Analyse auf den ersten Blick ineffizient erscheint, am Ende insgesamt doch effizienter ist.

analysis type	Sum Input	Avg Input	Sum Output	Avg Ouput	Sum Seconds	Avg Seconds
sentiment	643206	1072	269 479	449	3023.938	5
technical	6933314	11556	469 664	783	6336.445	11

Tabelle 4: Summe und Durchschnitt der verbrauchten Token und Zeit

Die Analyse der in Tabelle 3 dargestellten Ergebnisse verdeutlicht, dass die Stimmungsanalyse im Vergleich zur technischen Analyse signifikant effizienter ist. Die Untersuchung basiert auf einer Stichprobe bestehend aus jeweils 600 Stimmungs- und 600 technischen Analysen der Tabelle *actions*. Die Anzahl der verarbeiteten Input-Token beträgt in der Stimmungsanalyse 643.206, was weniger als 10 % der 6.933.314 Token der technischen Analyse entspricht. Auch der durchschnittliche Tokenverbrauch pro Stimmungsanalyse ist um mehr als den Faktor 10 geringer. Hinsichtlich der Output-Token ist der Unterschied weniger stark ausgeprägt: Während die Stimmungsanalyse 269.497 Token für 600 Analysen erzeugte, lag die Anzahl bei der technischen Analyse bei 469.664 Token. Die Gesamtberechnungszeit der technischen Analyse beläuft sich auf 6336 Sekunden (1,75 Stunden) und ist damit mehr als doppelt so hoch wie die der Stimmungsanalyse mit 3024 Sekunden (50 Minuten).

Die analysierten Daten zeigen eindeutig, dass die Stimmungsanalyse im Analyseschritt eine höhere Effizienz aufweist, da sowohl der Rechenaufwand als auch die Anzahl der verbrauchten Token reduziert werden, was zu niedrigeren Kosten führt. Die Kosten für eine Million Input-Token des Modells GPT-4o-mini betragen 0,15 US-Dollar, während eine Million Output-Token mit 0,60 US-Dollar zu veranschlagen sind. Demnach beliefen sich die Kosten der Stimmungsanalyse für 30 Unternehmen im Monat Februar 2024 auf 0,26 US-Dollar, während die technische Analyse Kosten in Höhe von 1,32 US-Dollar verursachte. Somit erweist sich die Stimmungsanalyse in diesem Experiment als etwa fünfmal kosteneffizienter als die technische Analyse.

Dieser erhebliche Unterschied lässt sich darauf zurückführen, dass in der Stimmungsanalyse ausschließlich Nachrichtenartikel berücksichtigt werden, die am aktuellen oder vorangegangenen Tag veröffentlicht wurden. Zudem werden pro Tool-Aufruf maximal 10 Artikel mit dem höchsten Relevanzscore aus allen verfügbaren Artikeln extrahiert, wodurch die Anzahl der verarbeiteten Input-Token minimiert wird. Insbesondere kleinere Aktiengesellschaften weisen häufig keine relevanten Nachrichtenartikel auf, sodass viele Artikel bereits im ersten Schritt der Stimmungsanalyse ausgeschlossen werden. Dies führt dazu, dass für diese Unternehmen an Tagen mit geringer Nachrichtenlage kaum Token verbraucht werden. In der technischen Analyse tritt dieser Effekt nicht auf, da für jedes Unternehmen – unabhängig von seiner Größe oder Medienpräsenz – die gleichen mathematisch berechneten Datenpunkte existieren und somit stets eine gleichbleibende Datenmenge verarbeitet wird.

Zudem ist zu berücksichtigen, dass das LLM bei der Stimmungsanalyse bereits vorverarbeitete Nachrichtenartikel mit zugehörigem Stimmungs- und Relevanzscore erhält. Daher muss es nicht erneut den gesamten Artikel analysieren, was sowohl die Verarbeitungszeit als auch den Tokenverbrauch erheblich reduziert. Dennoch ist zu beachten, dass im ersten Schritt der Stimmungsanalyse die ersten 600 Token eines jeden untersuchten Artikels verarbeitet wurden. Da im zweiten Schritt bis zu 10 Artikel berücksichtigt werden, würde sich der Tokenverbrauch pro Stimmungsanalyse um mindestens 6000 Token erhöhen. Darüber hinaus wurden im ersten Schritt der Analyse häufig mehr als 10 Artikel

gesichtet, sofern über die Plattformen GNews und Alphavantage eine größere Anzahl verfügbar war. Somit ist der tatsächliche Tokenverbrauch der Stimmungsanalyse pro Aktie und Tag tendenziell höher als in der direkten Gegenüberstellung mit der technischen Analyse ersichtlich.

analysis type	Sum Input	Avg Input	Sum Output	Avg Ouput	Sum Seconds	Avg Seconds
sentiment	142958	2383	41 336	689	362.343	6
technical	702109	11702	46 907	782	656.517	11

Tabelle 5: Summe und Durchschnitt der verwendeten Token und der verbrauchten Zeit von Apple, Microsoft, Tesla und Nvidia

Auffällig ist zudem, dass sich die Effizienzunterschiede zwischen den beiden Analysemethoden verringern, wenn ausschließlich die Aktien von Apple, Nvidia, Tesla und Microsoft betrachtet werden (Tabelle 4), also Unternehmen, die eine höhere Wahrscheinlichkeit für existierende Nachrichten haben. Insbesondere im Bereich der Output-Token gleichen sich die Werte beider Analysen weitgehend an. Die zeitliche Differenz ist um 1 Sekunden kleine geworden. Dies ist darauf zurückzuführen, dass diese Unternehmen eine durchgängig hohe Medienpräsenz aufweisen, wodurch an jedem Tag mindestens ein relevanter Nachrichtenartikel im Tool-Calling-Prozess zurückgegeben wird.

4.3 Praktikabilität

Die Praktikabilität der beiden Analysemethoden ist im Rahmen dieses Experiments als gleichwertig zu bewerten. Dies ist darauf zurückzuführen, dass in der Software wiederverwendbare Methoden implementiert wurden, die durch unterschiedliche Parameter verschiedene Funktionen ausführten. Daher spielte es keine wesentliche Rolle, ob die Stimmungsanalyse oder die technische Analyse durchgeführt wurde. Die Prompts beider Methoden waren schlicht gestaltet und enthielten lediglich das Datum sowie den Ticker der zu analysierenden Aktie. Für das Tool-Calling wurden jeweils spezifische JSON-Objekte erstellt, die hinsichtlich des Implementierungsaufwands vergleichbar waren. Die Datenerfassung aus der Datenbank wurde von ChatGPT eigenständig übernommen, nachdem die Services für die Datenbankabfragen erstellt wurden.

Ein Unterschied in der Implementierung bestand jedoch in der Erfassung der Ausgangsdaten. Da die Stimmungsanalyse im Rahmen dieser Untersuchung auf externe Nachrichtenquellen angewiesen war, musste ein zusätzlicher Zwischenschritt zur Extraktion der relevanten Inhalte aus den Webseiten implementiert werden. Dieser Schritt war für die technische Analyse nicht erforderlich. Allerdings erwies sich das Implementieren des Services zum Speichern der Nachrichtenartikel als weniger aufwändig als das Speichern der technischen Daten, wodurch sich der zeitliche Gesamtaufwand beider Methoden insgesamt annähernd ausglich.

Obwohl die Stimmungsanalyse einen zusätzlichen Prozessschritt in Form der Datenerhebung und anschließenden Bewertung durch ChatGPT erforderte, konnte dieser vollständig automatisiert innerhalb einer einzelnen API-Anfrage an die Software erfolgen. Im Gegensatz dazu war dies für die technische Analyse nicht umsetzbar, da die Abfrage der technischen Daten über Alphavantage aufgrund einer Limitierung auf 25 Anfragen pro Tag über mehrere Tage verteilt erfolgen musste.

4.4 Genauigkeit

Das letzte Bewertungskriterium umfasst die Analyse der Genauigkeit der Handelsempfehlungen. Hierzu werden die Gewinn-Verlust-Quote betrachtet, die durchschnittliche Rendite pro Trade berechnet und abschließend die Gesamtrendite aller Trades evaluiert. Darüber hinaus erfolgt ein Vergleich der Performance der verschiedenen Strategien zur Schließung der Trades. Es wurden insgesamt 3540 Handelsempfehlungen gesammelt und analysiert. Dafür wurden für alle untersuchten Unternehmen an jedem Handelstag der Monate Februar-, Juni- und September 2024 Handelsempfehlungen generiert.

4.4.1 Gewinn-Verlust-Quote

Die Gewinn-Verlust-Quote ist eine Kennzahl zur Bewertung der Profitabilität von Handelsstrategien. Sie wird berechnet, indem die durchschnittliche Rendite aller erfolgreichen Trades durch die durchschnittliche Rendite aller verlustreichen Trades dividiert wird. Ein Wert größer als 1 zeigt an, dass die Handelsstrategie insgesamt profitabel war, da die durchschnittlichen Gewinne die durchschnittlichen Verluste übersteigen.

Zur Berechnung dieser Kennzahl wurde eine Prozedur entwickelt, die ein Array aus Zeichenketten als Eingabeparameter verwendet. Diese Funktionalität ermöglicht es, Ausreißer aus der Analyse zu entfernen, indem bestimmte Unternehmen von der Untersuchung ausgeschlossen werden. Die Berechnung des durchschnittlichen Gewinns erfolgt durch eine Filterung der Daten auf Trades mit einem positiven Ergebnis (WIN). Analog dazu wird die durchschnittliche Verlustquote durch eine Selektion der verlustreichen Trades (LOSS) ermittelt. Um eine Division durch Null zu vermeiden, wird vor der Berechnung der Quote eine entsprechende Überprüfung durchgeführt.

Zusätzlich wird in der WHERE-Klausel sichergestellt, dass Unternehmen, deren Ticker im übergebenen Array enthalten sind, von der Analyse ausgeschlossen werden. Zur Gewährleistung einer differenzierten Auswertung erfolgt zudem eine Gruppierung nach relevanten Merkmalen, um für jede Gruppe individuelle Aggregationen durchführen zu können.

Analysis Type	Closing Strategy Name	Avg Win (USD)	Avg Loss (USD)	Profit Loss Ratio
technical	day_1_close	2.48	2.66	0.93
technical	day_2_close	3.07	3.47	0.88
technical	day_3_close	3.61	4.18	0.86
technical	day_close	1.67	1.48	1.13
technical	take_profit	5.01	2.64	1.89
sentiment	day_1_close	2.41	2.50	0.96
sentiment	day_2_close	3.03	3.48	0.87
sentiment	day_3_close	3.57	4.57	0.78
sentiment	day_close	1.40	1.34	1.05

Tabelle 6: Gewinn-Verlust-Quote der technischen Analyse und Stimmungsanalyse

Die Ergebnisse in Tabelle 5 belegen, dass die Gewinn-Verlust-Quote im Rahmen der Stimmungsanalyse ausschließlich dann positiv ausfällt, wenn ein Trade noch am selben Tag geschlossen wird, an dem er eröffnet wurde. Bereits ein Handelsschluss einen Tag nach Eröffnung führt – bei über 1770 betrachteten Trades – zu Verlusten. Dieses Resultat stimmt den Ergebnissen von Lopez-Lira und Tang [55] überein, die ebenfalls zeigen, dass die besten Ergebnisse erzielt werden, wenn Trades am gleichen Tag beendet

werden. Zwar wird von Lopez-Lira und Tang [55] postuliert, dass die Effekte von Nachrichten bis zu zwei Tage anhalten können, jedoch konnte dieser Effekt in der vorliegenden Analyse nicht bestätigt werden. Vielmehr nimmt die Gewinn-Verlust-Quote mit zunehmendem zeitlichen Abstand zum Eröffnungstag stetig ab. Zwar weist der Handelsschluss nach sechs Tagen eine Quote von 0.805 auf, was den 0.78 nach drei Tagen überlegen erscheint, doch handelt es sich hierbei vermutlich um statistische Zufälligkeiten, da die Werte für vier und fünf Tage mit 0.65 beziehungsweise 0.69 deutlich niedriger ausfallen.

Hinsichtlich der technischen Analyse zeigt sich, dass die Take-Profit-Strategie die höchste Gewinn-Verlust-Quote erzielt – sie liegt mit einem Wert von 1.89 signifikant über der besten Quote von 1.05, die im Rahmen der Stimmungsanalyse erreicht wurde. Als guter Wert wird eine Quote von circa 2 betrachtet [65], sodass die Technische Analyse diesem Ziel nahezu gerecht wird. Auffällig ist zudem, dass Trades, die auf Basis der Technischen Analyse durchgeführt werden, ebenfalls eine positive Quote aufweisen, sofern sie noch am Eröffnungstag geschlossen werden. Allerdings ist dieser Befund insofern weniger aussagekräftig, als die zugrunde liegenden Handelsempfehlungen von ChatGPT auf der Verwendung von Take-Profit- und Stop-Loss-Werten basieren und nicht zeitlicher Natur waren.

Analysis Type	Closing Strategy	Attribute Value	Avg Win (USD)	Avg Loss (USD)	Profit Loss Ratio
technical	take_profit	small	8.85	4.06	2.18
technical	take_profit	large	3.44	1.98	1.73
technical	take_profit	medium	4.72	2.21	2.13
sentiment	day_close	small	1.89	2.27	0.83
sentiment	day_close	medium	1.59	1.20	1.33
sentiment	day_close	large	1.29	1.06	1.22

Tabelle 7: Gewinn-Verlust-Quote gruppiert nach Unternehmensgröße

Wird die Gewinn-Verlust-Quote der Unternehmen nach ihrer Größe gruppiert betrachtet (Tabelle 6), so zeigt sich, dass die Quote der Stimmungsanalyse in der Teilmenge mittelgroßer und großer Unternehmen deutlich besser ausfällt als im Gesamtdurchschnitt. Konkret liegen die Quotienten für mittelgroße beziehungsweise große Unternehmen bei 1.33 bzw. 1.22.

Im Gegensatz dazu manifestiert sich diese Diskrepanz bei der technischen Analyse nicht in gleicher Weise. So weicht die Quote großer Unternehmen nur geringfügig vom Gesamtdurchschnitt ab, während kleinere und mittlere Unternehmen eine deutlich höhere Quote von über 2,1 aufweisen.

Analysis Type	Closing Strategy	Attribute Value	Avg Win (USD)	Avg Loss (USD)	Profit Loss
technical	take_profit	consumer goods and services	4.36	2.20	1.98
technical	take_profit	technology	5.49	2.62	2.09
technical	take_profit	healthcare	5.87	3.18	1.85
technical	take_profit	energy and industry	5.74	2.67	2.15
sentiment	day_close	technology	1.26	1.41	0.89
sentiment	day_close	healthcare	2.30	1.10	2.10
sentiment	day_close	energy and industry	1.99	1.37	1.45
sentiment	day_close	consumer goods and services	1.32	1.10	1.20

Tabelle 8: Gewinn-Verlust-Quote gruppiert nach Sektor

Im Rahmen der sektoralen Gruppierung (Tabelle 7) zeigt sich, dass die technische Analyse in den verschiedenen Branchen homogene Ergebnisse liefert. So liegt der niedrigste Gewinn-Verlust-Quotient im Gesundheitssektor bei 1,85, während im Bereich Energie und Industrie mit 2.15 der höchste Wert

erzielt wird. Demgegenüber zeichnet sich das Bild der Stimmungsanalyse durch eine stärkere Differenzierung aus: Hier dominiert der Gesundheitssektor mit einer Quote von 2.1, während der Technologiesektor mit lediglich 0.9 deutlich unter dem Durchschnitt liegt.

Analysis Type	Closing Strategy	Attribute Value	Avg Win (USD)	Avg Loss (USD)	Profit Loss Ratio
technical	take_profit	medium	3.44	1.72	2.00
technical	take_profit	high	7.26	3.47	2.09
technical	take_profit	low	2.16	1.77	1.22
sentiment	day_close	low	0.95	0.53	1.80
sentiment	day_close	medium	0.95	0.86	1.10
sentiment	day_close	high	2.11	1.80	1.17

Tabelle 9: Gewinn-Verlust-Quote gruppiert nach Volatilität

Bei der Gruppierung der Unternehmen nach Volatilität zeigt sich, dass die technische Analyse bei niedrig volatilen Unternehmen mit einem Gewinn-Verlust-Quotienten von 1.22 signifikant schlechter abschneidet als bei Unternehmen mit mittlerer oder hoher Volatilität, wo Quotienten von etwa 2 erzielt werden.

Im Gegensatz dazu erzielt die Stimmungsanalyse bei Unternehmen mit niedriger Volatilität überdurchschnittliche Ergebnisse, wobei hier ein Quotient von 1.8 beobachtet wird, während die Quotienten in den anderen Volatilitätsgruppen im Durchschnitt liegen.

4.4.2 Durchschnittliche Rendite pro Trade

Die durchschnittliche Rendite pro Trade wird ebenfalls durch eine Prozedur berechnet, die den Mittelwert aller Renditen für eine spezifische Schließungsstrategie ermittelt.

Analog zur Berechnung der Gewinn-Verlust-Quote ermöglicht die Implementierung eines Arrays als Eingabeparameter die gezielte Filterung bestimmter Unternehmen aus der Analyse. Falls keine Unternehmen ausgeschlossen werden, erfolgt die Berechnung der absoluten Rendite durch Multiplikation der prozentualen Rendite mit dem Startpreis des jeweiligen Trades. Anschließend wird der Durchschnitt dieser Werte über alle betrachteten Trades hinweg bestimmt, um eine präzise Einschätzung der durchschnittlichen Handelsrendite für die jeweilige Closing Strategie zu erhalten.

Analysis Type	Closing Strategy	Average Return	Average Percentage
technical	day_1_close	-0.24	0.14
technical	day_2_close	-0.76	0.02
technical	day_3_close	-1.32	0.00
technical	day_close	0.20	0.27
technical	take_profit	0.99	0.88
sentiment	day_1_close	-0.67	0.18
sentiment	day_2_close	-1.98	0.25
sentiment	day_3_close	-3.38	0.10
sentiment	day_close	0.22	0.10

Tabelle 10: Durchschnittliche Rendite Pro Trade

Die durchschnittliche Rendite der Stimmungsanalyse bei Tagesschluss beträgt lediglich 0,1 %, was die Aussagekraft über die Funktionalität dieser Strategie aufgrund des nahezu neutralen Ergebnisses einschränkt. Länger gehaltene Trades generieren ebenfalls keine positiven Ergebnisse, da sie negative Renditen aufweisen.

Im Gegensatz dazu erzielt die technische Analyse mit der Take-Profit-Strategie einen Durchschnitt von 0,88 % und erweist sich somit als leicht profitabel.

4.4.3 Gesamtrendite

Die Gesamtrendite aggregiert die Renditen aller Trades über die verschiedenen Schließungsstrategien hinweg. Ihre Berechnung folgt einem ähnlichen Ansatz wie die Bestimmung der durchschnittlichen Rendite pro Trade. Allerdings wird hier anstelle des Durchschnitts die Gesamtsumme aller Renditen ermittelt. Dies ermöglicht eine Einschätzung darüber, welchen absoluten Gewinn oder Verlust eine bestimmte Analyse- und Schließungsstrategie insgesamt generiert hätte, wenn davon ausgegangen wird, dass jedes Mal genau eine Aktie gekauft wird.

Um die Gesamtrendite in Prozent auszudrücken, wird die Summe aller absoluten Renditen durch die Summe der jeweiligen Eröffnungspreise der Trades dividiert. Dadurch wird eine relative Betrachtung der erzielten Rendite in Bezug auf das eingesetzte Kapital ermöglicht, wodurch sich die Effektivität der unterschiedlichen Analyse- und Schließungsstrategien besser vergleichen lässt.

Analysis Type	Closing Strategy	Overall Return	Overall Percentage
technical	day_1_close	-417.93	-0.24
technical	day_2_close	-1309.85	-0.75
technical	day_3_close	-2273.18	-1.29
technical	day_close	343.73	0.20
technical	take_profit	1699.60	0.97
<hr/>			
sentiment	day_1_close	-451.97	-0.42
sentiment	day_2_close	-1328.36	-1.23
sentiment	day_3_close	-2267.71	-2.11
sentiment	day_close	148.64	0.14

Tabelle 11: Gesamtrendite aller Trades

Die Analyse von Tabelle 10 zeigt, dass die Stimmungsanalyse bei Tagesschluss eine durchschnittliche Rendite von 0,14 % aufweist, was nahezu einem Nullergebnis entspricht. Alle anderen Strategien zur Schließung von Trades der Stimmungsanalyse generieren negative Renditen. Bei einer Haltedauer von drei Tagen resultiert bereits ein Verlust von 2,11 %. Dies deutet darauf hin, dass die Genauigkeit der Stimmungsanalyse mit zunehmender Haltedauer abnimmt.

Die technische Analyse erzielt eine Gesamtrendite von 0,97 %, was potenziell profitabel ist, solange die Transaktionskosten unter 0,97 % liegen. Auch hier ist der Tagesschluss mit 0,2 % profitabel, jedoch ist der Unterschied zur Stimmungsanalyse gering. Trades, die nach bis zu drei Tagen geschlossen werden, sind negativ. Interessanterweise sind die Verluste mit -1,29 % geringer als bei der Stimmungsanalyse.

Bei der Stimmungsanalyse als auch bei der technischen Analyse sind die Gesamtrenditen höher als der Durchschnitt pro Trade. Die Gesamtrendite könnte höher sein als die durchschnittliche Rendite pro Trade,

da sie die kumulative Wirkung aller Trades über einen bestimmten Zeitraum widerspiegelt und somit von der Anzahl der Trades und deren individuellen Volumina beeinflusst wird. Zudem können einzelne, besonders profitable Trades die Gesamtrendite stärker beeinflussen als der Durchschnitt aller Trades, weil der Durchschnitt alle Trades gleichwertig behandelt, während die Gesamtrendite die tatsächliche Summe der Gewinne und Verluste darstellt. Ein einzelner Trade mit hohem Gewinn kann einen größeren Einfluss auf diese Summe haben als viele kleine Verluste oder Gewinne, wodurch die Gesamtrendite unverhältnismäßig stark ansteigt.

Analysis Type	Closing Strategy	Attribute Value	Overall Return	Overall Percentage
technical	day_close	low	16.78	0.06
technical	day_close	high	250.55	0.25
technical	day_close	medium	76.41	0.15
technical	take_profit	medium	336.48	0.68
technical	take_profit	high	1510.25	1.52
technical	take_profit	low	-147.13	-0.56
sentiment	day_close	low	-4.92	-0.08
sentiment	day_close	medium	20.15	0.08
sentiment	day_close	high	133.41	0.18

Tabelle 12: Gesamtrendite gruppiert nach Volatilität

Tabelle 11 offenbart, dass die Stimmungsanalyse, gruppiert nach Unternehmensvolatilität, keine signifikanten Unterschiede aufweist. Unternehmen mit niedriger Volatilität weisen eine negative Gesamtrendite auf, während mittlere und hohe Volatilität mit 0,08 % bzw. 0,18 % kaum vom Gesamtdurchschnitt abweichen. Dennoch deutet sich an, dass eine hohe Volatilität tendenziell zu besseren Ergebnissen führt.

Die technische Analyse zeigt ebenfalls, dass Unternehmen mit hoher Volatilität besonders gut performen und eine Rendite von 1.5% haben. Unternehmen mit geringer Volatilität hingegen erzielen eine Rendite von -0,65 %. Die Ergebnisse legen nahe, dass die Effektivität der technischen Analyse mit zunehmender Unternehmensvolatilität steigt. Dies steht im Einklang mit der Gewinn-Verlust-Quote, die bei hoher Volatilität ebenfalls hoch war.

Analysis Type	Closing Strategy	Attribute Value	Overall Return	Overall Percentage
technical	take_profit	consumer goods and services	465.04	1.01
technical	take_profit	technology	1149.95	1.36
technical	take_profit	healthcare	-37.37	-0.18
technical	take_profit	energy and industry	121.98	0.51
sentiment	day_close	technology	69.54	0.10
sentiment	day_close	healthcare	2.22	0.06
sentiment	day_close	energy and industry	7.74	0.12
sentiment	day_close	consumer goods and services	69.15	0.23

Tabelle 13: Gesamtrendite gruppiert nach Sektor

Die sektorale Analyse der Stimmungsanalyse offenbart, dass Nachrichten im Bereich „Consumer Goods and Services“ einen signifikanten Einfluss haben, mit einer Rendite von 0,23 %, die 0,1 % über dem Gesamtdurchschnitt liegt. Alle übrigen Sektoren weisen unterdurchschnittliche Renditen auf.

Im Kontext der technischen Analyse ist der Technologiesektor mit einer Rendite von 1,36 % hervorzuheben, die deutlich über dem Durchschnitt liegt. Im Gegensatz dazu erweist sich die technische Analyse im Gesundheitssektor als wenig effektiv und generiert eine negative Rendite von -0,18 %.

4.4.4 Trefferquote

Die Trefferquote wird in Prozent angegeben und wird berechnet, indem die Anzahl der gewonnenen Trades durch die Anzahl aller Trades in der untersuchten Gruppe dividiert wird. Das Ergebnis ist in Tabelle 13 zu sehen.

Analysis Type	Closing Strategy	Win Count	Loss Count	Accuracy (%)
Technical	"day close"	953	765	55.47
Technical	"take profit"	790	926	46.03
Sentiment	"day close"	352	319	52.43

Tabelle 14: Trefferquote beider Analysemethoden über alle Unternehmen hinweg

Es lässt sich feststellen, dass die "Day-Close"-Strategie die höchste Trefferquote aufweist und somit die größte Genauigkeit erzielt. Dies deutet darauf hin, dass Handelsentscheidungen, die auf technischer Analyse basieren und am selben Tag geschlossen werden, die höchste Wahrscheinlichkeit aufweisen, profitabel zu sein. Auch die auf Stimmungsanalysen basierende Strategie zeigt mit einer Trefferquote von über 50 % eine positive Erfolgsbilanz. Im Gegensatz dazu weist die "Take-Profit"-Strategie mit einer Trefferquote von lediglich 46,03 % eine geringere Erfolgswahrscheinlichkeit auf und ist im Durchschnitt mit einem negativen Handelsergebnis verbunden.

Analysis Type	Strategy	Volatility	Win Count	Loss Count	Accuracy (%)
Technical	take_profit	low	55	120	31.43
Technical	take_profit	medium	322	360	47.21
Technical	take_profit	high	413	446	48.08
Sentiment	day_close	low	17	24	41.46
Sentiment	day_close	medium	138	116	54.33
Sentiment	day_close	high	197	179	52.39

Tabelle 15: Trefferquote gruppiert nach Volatilität

Analysis Type	Strategy	Size	Win Count	Loss Count	Accuracy (%)
Technical	take_profit	small	180	272	39.82
Technical	take_profit	medium	203	254	44.42
Technical	take_profit	large	407	400	50.43
Sentiment	day_close	small	22	17	56.41
Sentiment	day_close	medium	97	71	57.74
Sentiment	day_close	large	233	231	50.22

Tabelle 16: Trefferquote gruppiert nach Unternehmensgröße

Auffällig ist, dass bei der technischen Analyse ein klarer Trend zu sehen ist. Je größer und volatiler das untersuchte Unternehmen ist, desto besser ist die Trefferquote. Bei der Stimmungsanalyse ist nicht direkt ein klares Muster zu erkennen, da die Trefferquote bei mittlerer Volatilität und Größe am höchsten ist. Die Trefferquote der technischen Analyse übersteigt bei großen Unternehmen die 50 % Mark. Die Genauigkeit ist bei hochvolatilen Unternehmen am höchsten.

4.4.5 Interpretation der Genauigkeit

Die Beobachtung, dass die Stimmungsanalyse bei kleinen Unternehmen eine geringere Rendite aufweist, könnte auf die geringere Informationsdichte zurückzuführen sein. Kleine Unternehmen generieren tendenziell weniger medienwirksame Nachrichten, was zu einer unvollständigen oder verzerrten Stimmungsbewertung führen kann. In der Datenbank werden nur Artikel gespeichert, die laut Analyse durch ChatGPT eine hohe Relevanz haben. Kleinere Unternehmen haben dadurch an vielen Tagen keinen einzigen Artikel zur Verfügung.

Die überdurchschnittliche Performance der Stimmungsanalyse im Gesundheitssektor könnte auf die Art der Nachrichten in diesem Sektor zurückzuführen sein. Ereignisse wie die Zulassung neuer Medikamente oder bahnbrechende Forschungsergebnisse generieren oft klare und prägnante Nachrichten, die sich direkt auf die Aktienkurse auswirken. Im Vergleich dazu sind Nachrichten im Technologiesektor möglicherweise weniger eindeutig, was die Stimmungsbewertung erschwert.

Geringere Volatilität führte bei der Stimmungsanalyse zu einer besseren Gewinn-Verlust-Quote. Dies könnte auf stabilere Fundamentaldaten und eine weniger störende Nachrichtenlage zurückzuführen sein. In solchen Fällen werden die aus Nachrichten abgeleiteten Stimmungswerte klarer und konsistenter, da kurzfristige Marktschwankungen und spekulative Effekte, die bei volatileren Unternehmen stärker zum Tragen kommen, hier weitgehend ausbleiben. Somit ermöglicht die geringere "Rauschkomponente" in den Daten eine präzisere und zuverlässigere Empfehlung.

Die überdurchschnittliche Performance der technischen Analyse bei Unternehmen mit hoher Volatilität lässt sich durch den zugrundeliegenden Ansatz des Trendtradings erklären. Trendtrading erweist sich insbesondere bei stark volatilen Unternehmen als erfolgreich, da deren ausgeprägte Kursbewegungen die Identifikation und Nutzung von Trends erleichtern. Diese Beobachtung steht im Einklang mit der Fachliteratur, die argumentiert, dass hohe Volatilität zu größeren Preissprüngen führt, wodurch Trendfolge-Strategien besonders profitabel sind. [38,51]

Dass die technische Analyse im Technologiesektor besser funktioniert als im Gesundheitssektor, könnte mit der Volatilität innerhalb der Sektoren zusammenhängen. Die vorliegende Unternehmensauswahl im Technologiesektor besteht zu einem Großteil aus Unternehmen mit hoher Volatilität, während im Gesundheitssektor ein höherer Anteil an Unternehmen mit geringerer Volatilität vertreten ist.

Es ist von Bedeutung festzustellen, dass die Take-Profit-Strategie im Rahmen der technischen Analyse die höchste Rendite erzielt, obwohl sie die niedrigste Trefferquote aufweist. Dies könnte darauf zurückzuführen sein, dass bei korrekten Empfehlungen die hohen Zielwerte der Take-Profit-Marke zu überdurchschnittlich hohen Gewinnen geführt haben und dass durch die Stop-Loss Marke die Verluste im Rahmen gehalten werden konnten.

Die Tatsache, dass die technische Analyse lediglich bei großen Unternehmen eine positive Trefferquote aufweist, lässt sich möglicherweise dadurch erklären, dass sie auf dem Prinzip der selbsterfüllenden Prophezeiung basiert. Da bei großen Unternehmen eine größere Anzahl von Marktteilnehmern die technische Analyse anwendet, entfaltet sie dort eine stärkere Wirkung und erzielt entsprechend bessere Ergebnisse als bei kleinen Unternehmen.

Zusammenfassend lässt sich festhalten, dass die Genauigkeit der Handelsempfehlungen sowohl von der Art der Analyse, als auch von verschiedenen Einflussfaktoren wie Unternehmensgröße, Sektor und Volatilität abhängt. Die vorliegenden Ergebnisse liefern wertvolle Einblicke in die Stärken und

Schwächen der verschiedenen Ansätze und können dazu beitragen, Handelsstrategien zu optimieren und das Risikomanagement zu verbessern. Es ist zu sehen, dass die vorliegende Untersuchung die Volatilität und den Sektor des Unternehmens als wesentliche Einflussfaktoren auf die Genauigkeit der Handelsempfehlungen identifiziert und dass die technische Analyse bessere Renditen liefert als die Stimmungsanalyse, obwohl die Trefferquote insgesamt schlechter ist. Unabhängig davon ist für die technische Analyse die Take Profit Strategie am besten geeignet – die Stimmungsanalyse hatte mit der Schließung der Trades am Tag ihrer Öffnung die besten Ergebnisse.

5 Ergebnis

Dieses Kapitel widmet sich der kritischen Bewertung der Ergebnisse. Dabei werden zunächst die Vor- und Nachteile der beiden untersuchten Analysemethoden detailliert herausgearbeitet. Anschließend werden Beobachtungen und Erkenntnisse aus der praktischen Arbeit mit ChatGPT und dessen API dargestellt, gefolgt von dem daraus abgeleiteten Kapitel "Lessons Learned".

5.1 Technische Analyse

Die technische Analyse konnte eine nahezu siebenmal höhere Gesamtrendite über 1.770 Trades erzielen als die Stimmungsanalyse. Dieses Ergebnis verdeutlicht das Potenzial der technischen Analyse. Dennoch ist es wichtig, die Vor- und Nachteile dieser Methode, insbesondere in Verbindung mit der Nutzung von ChatGPT, differenziert zu betrachten. Die folgenden Kapitel widmen sich einer Analyse der Stärken und Schwächen dieser Analysemethode.

5.1.1 Schwächen der technischen Analyse

In der Untersuchung wurde festgestellt, dass die technische Analyse mit zunehmendem Umfang des erlaubten Kontextes signifikant kostenintensiver wird. Konkret wurden in diesem Fall Finanzdaten der letzten 30 Tage sowie die Kursentwicklung der letzten 90 Tage herangezogen, um Muster im Aktienkurs zu identifizieren. Diese Herangehensweise ermöglichte es dem Modell, flexibel auf die Daten zuzugreifen, die es als für die Analyse relevant erachtete. Allerdings führte diese Freiheit dazu, dass eine erhebliche Anzahl an Token benötigt wurde, um eine konsistente Einschätzung zu generieren. Eine Erweiterung des Kontextfensters, beispielsweise auf drei Monate oder ein ganzes Jahr, würde die Kosten für eine einzelne Analyse weiter erheblich steigern.

Ein weiteres Problem ergibt sich aus der fehlenden wissenschaftlichen Evidenz für eine universell profitable Handelsstrategie. Es existieren keine eindeutigen und fundierten Ergebnisse, die belegen, dass eine bestimmte Methode der technischen Analyse dauerhaft erfolgreich ist. Dies erschwert die Bewertung der Ergebnisse der Untersuchung, da unklar bleibt, ob eine alternative Strategie möglicherweise zu besseren Resultaten führen würde.

Darüber hinaus zeigte sich eine Inkonsistenz in den generierten Handelsempfehlungen. Trotz der Verwendung identischer Daten und identischer Prompts kam es zu signifikanten Abweichungen in den Empfehlungen des Modells. Diese Variabilität stellt eine weitere Herausforderung für die Verlässlichkeit der technischen Analyse dar.

Außerdem ist es bei der technischen Analyse von hoher Relevanz die am besten funktionierenden Unternehmen herauszusuchen, da die Genauigkeit der Vorhersagen und Rendite in diesem Versuch stark von der Volatilität und der Größe des Unternehmens abhängt.

Ein zusätzlicher praktischer Nachteil war der Aufwand, der mit der Abfrage der technischen Daten über die API von Alphavantage verbunden war. Aufgrund eines täglichen Limits für Anfragen mussten die Daten über mehrere Tage hinweg gesammelt werden, was den Analyseprozess erheblich verzögerte.

Zusammenfassend zeigt die Untersuchung, dass die technische Analyse, insbesondere bei der Nutzung großer Datenmengen und flexibler Modelle, mit hohen Kosten, methodischen Unsicherheiten und praktischen Herausforderungen verbunden ist. Diese Limitationen sollten bei der Planung und Durchführung zukünftiger Nutzung berücksichtigt werden.

5.1.2 Stärken der technischen Analyse

Die Stärke der technischen Analyse liegt eindeutig in der Verfügbarkeit der Daten. Für jedes Unternehmen stehen sämtliche relevanten Kennzahlen zur Verfügung, da diese mathematisch berechnet werden. Dadurch können für jeden Handelstag qualitativ hochwertige technische Daten generiert und genutzt werden. Auch wenn das Sammeln und Speichern dieser Daten in der Datenbank, wie bereits bei den Schwächen erwähnt, mit einem hohen Aufwand verbunden ist, ist dies Besonderen dann der Fall, wenn kein Abonnement für die Datenabfrage vorhanden ist. Mit einem kostenpflichtigen Zugang zu externen APIs gestaltet sich das Sammeln der Daten deutlich einfacher und effizienter. Ein weiterer Vorteil der technischen Analyse ist, dass die Daten ohne zusätzliche Bearbeitung direkt in die Analyse einfließen können. Im Gegensatz zur Stimmungsanalyse, bei der Daten zunächst auf Relevanz oder Stimmungswerte untersucht und bewertet werden müssen, waren die technischen Daten sofort nutzbar.

Ein weiterer zentraler Vorteil der technischen Analyse ist die Möglichkeit, Verluste effektiv einzugrenzen. Durch die Implementierung von Stop-Loss-Mechanismen konnte bei den meisten Trades ein maximaler Verlust von lediglich 1–2 % realisiert werden. Dies trug wesentlich dazu bei, das Risiko zu minimieren und die Stabilität der Handelsstrategie zu gewährleisten. Darüber hinaus zeigte die technische Analyse bereits ohne eine gezielte Optimierung der ausgewählten Unternehmen eine positive Rendite. Zudem wurde beobachtet, dass die Rendite je nach Volatilität und Sektor der analysierten Unternehmen weiter gesteigert werden konnte. Dies deutet darauf hin, dass durch eine optimierte Auswahl der Unternehmen und eine gezielte Anpassung der Strategie noch bessere Ergebnisse erzielt werden könnten.

Der Nachteil, dass die zu untersuchenden Unternehmen ausgewählt werden müssen, um gute Ergebnisse zu erzielen, kann gleichzeitig auch als Vorteil gesehen werden. Es wurde deutlich, dass die richtige Auswahl an großen und volatilen Unternehmen die Rendite und Trefferquote steigerte. So kann das Ergebnis optimiert werden.

Zusammenfassend zeigt sich, dass die technische Analyse durch ihre breite Verfügbarkeit an Daten, die einfache Nutzbarkeit dieser Daten, die Möglichkeit zur Verlustbegrenzung und das Potenzial für weitere Optimierungen eine äußerst vielversprechende Methode darstellt.

5.2 Stimmungsanalyse

Obwohl die Stimmungsanalyse in diesem Fall kaum Rendite erbracht hat, sind dennoch spezifische Stärken und Schwächen dieser Analysemethode deutlich geworden, die in diesem Anwendungsfall zutage traten.

5.2.1 Schwächen der Stimmungsanalyse

Neben der offensichtlichen Feststellung, dass diese Analysemethode lediglich eine Gesamtrendite von 0,14 erzielt hat, zeigt sich ein weiterer Schwachpunkt darin, dass die Rendite auch durch eine

Differenzierung nach Volatilität, Sektor oder Firmengröße nicht signifikant verbessert werden konnte. Dies könnte darauf hindeuten, dass die Stimmungsanalyse auf Basis der verfügbaren Artikel nur begrenzt optimierbar ist.

In der Fachliteratur wird auch auf soziale Netzwerke wie Twitter zurückgegriffen [7,40], um Stimmungen zu analysieren. Dies war jedoch in dieser Untersuchung nicht möglich, da der Antrag auf Zugriff zur Twitter-API abgelehnt wurde. Die Ablehnung wurde damit begründet, dass die geplante Datennutzung nicht mit Artikel 34 des Digital Services Act übereinstimmt, da sie nicht der Identifizierung und Erkennung systemischer Risiken in der EU dient. Folglich konnten Twitter-Daten in dieser Analyse nicht berücksichtigt werden, was möglicherweise zu besseren Ergebnissen geführt hätte.

Ein weiteres Problem betrifft kleine Unternehmen, die nur eine geringe Medienpräsenz aufweisen. Für diese Unternehmen ist es schwierig, eine Stimmungsanalyse durchzuführen, da an vielen Tagen kaum relevante Nachrichtenartikel verfügbar sind. Dies erschwert es, fundierte Empfehlungen abzugeben, wenn nur wenige oder gar keine Informationen vorliegen. Im Gegensatz dazu sind technische Daten in solchen Fällen oft besser verfügbar. Es zeigt sich also, dass es eine Herausforderung darstellt, qualitativ hochwertige und konsistente Daten für die Stimmungsanalyse zu erhalten.

Darüber hinaus erfordert die Stimmungsanalyse einen hohen Ressourcenaufwand. Wie bereits erläutert, erfolgt die Analyse in zwei Schritten: Zunächst werden die Artikel analysiert und bewertet, bevor daraus eine Einschätzung abgeleitet wird, ob eine Aktie gekauft oder verkauft werden sollte. Für jeden Artikel mussten zahlreiche Token verwendet werden, da die Zusammenfassungen und die daraus resultierenden Scores andernfalls nicht präzise genug waren. Wenn sich jedoch herausstellte, dass ein Artikel irrelevant war, waren die dafür aufgewendeten Token verschwendet, da er nicht zur Erstellung einer Handelsanalyse genutzt werden kann.

5.2.2 Stärken der Stimmungsanalyse

Mit ausreichendem Kontext pro Artikel erwiesen sich die von ChatGPT berechneten Relevanz- und Stimmungsscores als präzise.

Die Stimmungsanalyse weist bei nahezu allen Unternehmensarten eine Trefferquote von über 50 % auf, was darauf hindeutet, dass sie im Durchschnitt effektiv ist. Allerdings muss hierbei berücksichtigt werden, dass die erzielte Rendite trotz der vergleichsweise hohen Genauigkeit nahezu bei null liegt.

Allerdings zeigte diese Untersuchung, dass die Stärken der Stimmungsanalyse insgesamt begrenzt sind. Insbesondere konnten keine weiteren signifikanten Vorteile oder Mehrwerte identifiziert werden, die über die punktuelle Genauigkeit der Scores hinausgehen. Dies deutet darauf hin, dass die Methode in diesem spezifischen Anwendungsfall nur eingeschränkt nutzbar ist.

5.3 Lessons Learned: Arbeiten mit ChatGPT

Im Rahmen der Arbeit mit der ChatGPT API sind zahlreiche Aspekte identifiziert worden, die zu schnelleren und besseren Ergebnissen in zukünftigen Anwendungen führen können, unabhängig vom jeweiligen Einsatzgebiet. Diese Erkenntnisse werden in diesem Kapitel als „Lessons Learned“ zusammengefasst.

5.3.1 Strukturierte Outputs

Die Nutzung strukturierter Outputs erwies sich als äußerst vorteilhaft, da die Antworten von ChatGPT stets dem vorgegebenen Format entsprachen und somit ein automatisiertes Verarbeiten der Ergebnisse ermöglichten. Dies erleichterte das zügige Durchlaufen großer Datenmengen und die Speicherung in der Datenbank erheblich. Allerdings führt die explizite Formatvorgabe auch zu Einschränkungen in der Antwortfreiheit, weshalb es sinnvoll erscheint, zusätzliche Felder zur Erklärung der Antworten bereitzustellen.

Diese Vorgehensweise bietet zwei zentrale Vorteile: Zum einen erhält das Modell einen definierten Bereich, in dem es sein Reasoning darlegen kann, sofern in der Prompt die Anforderung an Begründungen formuliert wurde – mit der Intention, die Performance zu verbessern. Zum anderen erleichtert die Möglichkeit gegebene Entscheidungen zu erläutern das Debugging erheblich. So wurde beispielsweise in den ersten Tests des Datenauslesens festgestellt, dass IDs zurückgegeben wurden, die nicht existierten. Durch die Ergänzung eines Erklärungsfeldes konnte jedoch schnell nachvollzogen werden, dass schlichtweg keine entsprechenden Daten verfügbar waren und ChatGPT sich daher alternative IDs ausgedacht hatte, um das vorgegebene Format einzuhalten. Dadurch konnten schnell Maßnahmen getroffen werden, wodurch dieses Problem behoben werden konnte.

Zudem ist es unerlässlich, bei der Verwendung strukturierter Outputs eindeutige und präzise Feldnamen für die JSON-Objekte zu wählen und jedem Feld eine klare Beschreibung beizufügen. So wurde im Beispiel zur Beschreibung des ID-Felds zunächst der Text „*The id of the dataobject that refers to the reasoning of this array item.*“ verwendet, was dazu führte, dass in den Antworten nicht die an das Modell übergebenen, sondern stattdessen erfundene bzw. unpassende IDs auftauchten. Durch die Anpassung der Beschreibung zu „*The id of the returned data from the functions that you use for the reasoning. For example if the get_price_data function returns items with id 100, 499 or 2 and you use the data with id 499 for the reasoning, this id should be put into here.*“ konnte mittels One-Shot-Prompting exemplarisch verdeutlicht werden, welche Daten in das jeweilige Feld einzutragen sind.

5.3.2 Tool Calling

Die Integration von Tool Calls – also die Bereitstellung eines Schemas mit Funktionsbeschreibungen und das Aufrufen dieser Funktionen auf Anforderung von ChatGPT – hat sich als relativ unkompliziert erwiesen. Dennoch konnten Erkenntnisse gewonnen werden, die beim Implementieren im nächsten Einsatz direkt einfließen können.

Eine präzise Formulierung der Prompts sowie der dazugehörigen Funktionsbeschreibungen ist dabei unabdingbar. Im Rahmen der Nutzung wurde explizit darauf geachtet, anzugeben, dass von den drei übergebenen Funktionen, abhängig von der gewählten Analysemethode, unterschiedliche Funktionen aufgerufen werden dürfen. Trotz der Vorgabe, dass ausschließlich Preis- und Nachrichtenartikel abgefragt werden sollten, forderte ChatGPT mitunter technische Daten an, obwohl dies nicht spezifiziert war. Das Modell nutzte demzufolge die Möglichkeit der übergebenen Tools, um diese Daten zu erhalten, was zu unerwünschten Ergebnissen führte. Daher erscheint es sinnvoll, in den übergebenen Tools ausschließlich diejenigen Funktionen zu integrieren, die im jeweiligen Aufruf zugelassen sind. Im vorliegenden Experiment wurden deshalb zwei Varianten bereitgestellt: technicalTools und sentimentalTools.

Ein weiterer Aspekt, der berücksichtigt werden sollte, ist die Überprüfung der Parameter, mit denen ChatGPT die Funktionen aufruft. In dem Experiment wurde häufig mit Start- und Enddatum gearbeitet, um den verfügbaren Datensatz einzuschränken. Obwohl in der Prompt vermittelt wurde, dass nur Daten innerhalb eines festen Zeitraums zulässig seien, kam es dennoch zu Aufrufen, bei denen ChatGPT diesen Rahmen nicht einhielt. Es erweist sich daher als zweckmäßig, die Parameter in der aufgerufenen Funktion anzupassen oder alternativ eine Fehlermeldung an ChatGPT zu übermitteln, welche auf die Ungültigkeit der Parameter hinweist.

Zudem hat sich gezeigt, dass es ratsam ist, eine maximale Anzahl an Antworten pro Durchlauf festzulegen. Da ChatGPT Funktionsaufrufe über das Flag „tool_call“ signalisiert, wurde in diesem Experiment die Anzahl solcher Anfragen auf zwei begrenzt, um sicherzustellen, dass die dritte Antwort stets die finale Antwort darstellt. Ohne diese Einschränkung wurden häufig mehr als drei Tool-Call-Antworten generiert, was zu Verzögerungen in der Antwortausgabe führte.

Schließlich wurde festgestellt, dass, selbst wenn in den Funktionsbeschreibungen sämtliche technischen Daten – etwa durch die Verwendung eines Enums, das alle möglichen Optionen definiert – aufgeführt werden, ChatGPT nicht alle diese Daten abruft. Die Ursachen hierfür können nicht endgültig geklärt werden. Es erscheint jedoch möglich, dass das Modell die in der Prompt explizit geforderten technischen Daten, wie beispielsweise ADX, aber auch nicht explizit genannte Daten wie RSI im Kontext der untersuchten Trades als irrelevant einstuft.

5.3.3 ChatGPT – API

Die API, über die mit dem LLM gearbeitet werden kann, erweist sich für kleinere Entwickler als wenig benutzerfreundlich, da sie durch strikte Rate-Limits erheblich eingeschränkt ist. Beispielsweise ist es mit dem Modell gpt-4o-mini lediglich möglich, bis zu 200.000 Token pro Minute zu übermitteln, während für gpt-4o nur 30.000 Token pro Minute zulässig sind. Diese Beschränkungen beruhen auf unterschiedlichen Preisstufen (Tiers): Ab einer Einzahlung von 5 USD befindet man sich in Tier 1, welches in diesem Experiment verwendet wurde. Erst mit einer Einzahlung von 50 USD erhöht sich das Rate-Limit für gpt-4o-mini auf 2.000.000 Token pro Minute. In vielen Anwendungsfällen stellt dies kein Problem dar, da typischerweise nicht 30 Anfragen gleichzeitig mit jeweils 10.000 Token gesendet werden; im vorliegenden Experiment war dies jedoch der Fall.

Um zu vermeiden, dass das Rate-Limit überschritten wird und Fehler auftreten, empfiehlt es sich, Anfragen in Batches aufzuteilen, die nacheinander in einer Queue abgearbeitet werden. Dadurch wird sichergestellt, dass nicht zu viele Anfragen parallel übermittelt werden. Die Batch-Größe sollte dabei in Abhängigkeit von der Anzahl der pro Element verwendeten Token gewählt werden – je weniger Token pro Element genutzt werden, desto mehr Anfragen können parallel bearbeitet werden.

Der Kostenaspekt der Anfragen darf ebenfalls nicht unterschätzt werden, auch wenn eine Million Input-Token bei gpt-4o-mini nur 0,15 USD kosten. In bestimmten Aufgabenbereichen, wie in diesem Experiment, können sehr viele Input-Token anfallen. Es ist zudem zu beachten, dass insbesondere die Output-Token meist um ein Vielfaches teurer sind (bei gpt-4o-mini etwa fünfmal, bei gpt-4o etwa viermal so hoch). Daher sollte im Vorfeld genau überlegt werden, welche Daten dem Modell tatsächlich übermittelt werden und wie umfangreich die Antwort ausfallen soll. Es empfiehlt sich, eine Kostenkalkulation auf Basis der durchschnittlichen Input- und Output-Token pro Anfrage vorzunehmen. Zudem kann es sinnvoll sein, Tests zunächst in einem kostengünstigeren Modell durchzuführen, das zwar

möglicherweise nicht die gleiche Performance wie das Zielmodell aufweist, jedoch ausreichend dazu dient, die Funktionsweise von Tool Calls oder strukturierten Outputs zu überprüfen – vorausgesetzt, beide Modelle unterstützen dieselben Funktionalitäten.

Ein weiterer Aspekt betrifft die Integration von Internet-Suchfunktionen. Keines der in der API verfügbaren Modelle unterstützt per Stand 10. Februar 2025 eine Internet-Suche. Dies bedeutet, dass es keinen Nutzen bringt, dem Modell externe Links zur Durchsicht zu übermitteln, da das Modell damit nicht arbeiten kann. Ein in diesem Experiment durchgeföhrter Versuch ergab zwar eine Antwort, diese basierte jedoch ausschließlich auf den im Link enthaltenen Informationen sowie dem im Modell integrierten Wissen. Daher ist es notwendig, Webscraping eigenständig zu implementieren, um aktuelle Daten zu erhalten.

6 Zusammenfassung

Die vorliegende Arbeit untersucht den Erfolg der technischen Analyse und der Stimmungsanalyse, wenn diese mithilfe von ChatGPT durchgeführt werden. Ziel der Untersuchung war es, sowohl die Stärken als auch die Schwächen beider Ansätze zu evaluieren und aufzuzeigen, inwiefern ChatGPT als unterstützendes Instrument zur Prognose von Aktienkursentwicklungen eingesetzt werden kann.

Im Rahmen der Arbeit wurde zunächst die Bedeutung eines präzisen Prompt-Engineerings hervorgehoben. Es zeigte sich, dass eine kluge Formulierung der Prompts, insbesondere unter Anwendung von *Few-Shot Prompting*, zu signifikant besseren Ergebnissen führen kann, ohne dass damit ein erheblicher Mehraufwand verbunden ist. Durch exemplarische Vorgaben gelang es dem LLM, den Kontext besser zu erfassen und sich an den übermittelten Beispielen zu orientieren. Ergänzend dazu wurde mittels *Instruction Prompting* in der DevPrompt spezifiziert, dass das Modell als Finanzexperte aufzutreten hat, sich je nach Analyseansatz auf die technische Analyse oder die Stimmungsanalyse zu konzentrieren hat und seine Ergebnisse durch *Reasoning* begründen soll. Darüber hinaus wurden im Prompt auch Erkenntnisse aus der Literatur zur technischen Analyse, beispielsweise unter Einbeziehung von Moving Averages Crossovers, integriert.

Die Datengrundlage der Analyse wurde eingehend vorgestellt. Es zeigte sich, dass die ChatGPT-API aufgrund ihrer beschränkten Fähigkeit, das Internet zu durchsuchen, ungenaue technische Daten lieferte, weshalb stattdessen auf Alphavantage zurückgegriffen wurde. Alphavantage lieferte korrekte und für jeden Handelstag verfügbare Daten. Zwar bietet der zugehörige Nachrichtenpunkt Relevanz- und Stimmungsscores, jedoch erwiesen sich diese Werte als nicht präzise genug. Daher wurden die entsprechenden Scores durch eigenes Crawling und anschließender Stimmungs- und Relevanzbewertung mittels ChatGPT generiert. Um Quelleninvarianz von Alphavantage auszugleichen, kam ergänzend die API von GNews zum Einsatz. Sämtliche gewonnenen Daten wurden anschließend in einer Datenbank zur weiteren Analyse gespeichert.

Zur praktischen Umsetzung wurde eine Softwarelösung entwickelt, die externe Finanzdaten von Alphavantage sowie Nachrichtenartikel aus unterschiedlichen Quellen bezieht. Diese Daten werden strukturiert an die ChatGPT-API übermittelt, wobei das LLM zur Erstellung von Handelsempfehlungen aufgefordert wird. Die Softwarearchitektur, einschließlich Datenbankstruktur und Aufbau der Applikation, wurde detailliert beschrieben. Die generierten Empfehlungen wurden abschließend mit verschiedenen simulierten Trading-Strategien verglichen. Hierbei kamen strukturierte Outputs zum Einsatz, um die umfangreiche Anzahl an Anfragen systematisch verarbeiten zu können, während Tool Calling es dem LLM ermöglichte, selbst zu entscheiden, welche Daten für eine Aktie an einem bestimmten Datum am besten geeignet sind.

Die Analyse berücksichtigte verschiedene Bewertungskriterien wie Konsistenz, Effizienz, Praktikabilität und Genauigkeit. Es zeigte sich, dass keine der beiden Methoden durchweg konsistente Ergebnisse lieferte, da bei identischen Inputs abweichende Antworten generiert wurden. Hinsichtlich der Effizienz schnitt die technische Analyse im eigentlichen Analyseschritt schlechter ab, da sie im Durchschnitt etwa sechs Sekunden länger benötigte, mehr als zehnmal so viele Input-Token und nahezu doppelt so viele Output-Token verwendete. Berücksichtigt man jedoch, dass der Stimmungsanalyse bereits im Vorfeld für jeden Artikel mindestens 600 Token als Input zugeführt wurden – wobei viele Artikel letztlich als irrelevant eingestuft wurden –, erweist sich die technische Analyse insgesamt als effizienter.

Beide Methoden erwiesen sich in Hinblick auf die Praktikabilität als gleichwertig, da sich bei korrekter Implementierung der entsprechenden Prompts und Strukturvorgaben kaum Unterschiede in der Erstellung und Auswertung der Analysen zeigten. Auffällig wurde jedoch der Unterschied in der Genauigkeit. So erwies sich für die technische Analyse die Take-Profit-Strategie als am profitabelsten, während bei der Stimmungsanalyse die Day-Close-Strategie – also das Schließen des Trades am Ende des Handelstages – den besten Ertrag lieferte. Keine der beiden Analyseformen überzeugte jedoch vollständig: Die Stimmungsanalyse erzielte im Durchschnitt über drei Monate lediglich eine vernachlässigbare Gesamtrendite von 0,14 %, während die technische Analyse geringfügig bessere Ergebnisse mit einer Gesamtrendite von 0,97 % aufwies. Durch Clusteranalysen der untersuchten Unternehmen ließ sich zudem feststellen, dass sowohl Volatilität, die Unternehmensgröße als auch der Tätigkeitssektor einen Einfluss auf die Performance der Analysemethoden haben. So erwies sich, dass bei Unternehmen hoher Volatilität die Take-Profit-Strategie der technischen Analyse eine Gesamtrendite von 1,52 % erzielte, wohingegen Unternehmen mit niedriger Volatilität lediglich eine Gesamtrendite von 0,68 % verzeichneten. Hinsichtlich der Stimmungsanalyse zeigte sich bei Unternehmen im Bereich *Consumer Goods and Services* eine geringfügig bessere Performance, wenngleich die erzielten Renditen von 0,23 % kaum über dem Durchschnitt lagen. Zudem wurde festgestellt, dass die Stimmungsanalyse bei kleineren Unternehmen mit eingeschränkter Medienpräsenz nur begrenzt einsetzbar ist, da an vielen Handelstagen nicht genügend relevante Nachrichtenartikel verfügbar waren.

Zusammenfassend lässt sich feststellen, dass ChatGPT in der untersuchten Anwendung nur begrenzt präzise und konsistente Vorhersagen der Aktienkursentwicklung liefert. Im Hinblick auf einen praktischen Einsatz als unterstützendes Prognoseinstrument erweist sich die technische Analyse als der Stimmungsanalyse vorzuziehen, da sie auf einer stabileren und umfassenderen Datenbasis basiert, während die Verfügbarkeit relevanter Medienartikel oftmals unzureichend ist. Dennoch kann eine innerhalb weniger Sekunden generierte Ersteinschätzung mit einem geringen Preis als ergänzende Unterstützung von Nutzen sein.

7 Diskussion und Ausblick

Diese Arbeit widmet sich einem dynamischen und hochaktuellen Forschungsfeld, das an der Schnittstelle von Künstlicher Intelligenz, Finanzmärkten und maschinengestützter Entscheidungsfindung liegt. Die Neuartigkeit des Themas bringt nicht nur spannende Möglichkeiten, sondern auch Herausforderungen mit sich, die im Folgenden kritisch reflektiert werden. Zudem wird ein Ausblick auf potenzielle Entwicklungen und zukünftige Forschungsrichtungen gegeben.

Eine der zentralen Herausforderungen ist die rasante Entwicklung der Technologie. Die kurze Gültigkeit von Erkenntnissen im Bereich der KI führt dazu, dass Ergebnisse bereits nach kurzer Zeit veraltet sein können. Ein Vergleich von Experimenten, die mit GPT-3 durchgeführt wurden, mit solchen, die auf GPT-4 basieren, verdeutlicht, wie stark sich die Leistungsfähigkeit von Modellen innerhalb weniger Monate verändern kann. Dieses Phänomen zeigte sich auch bei der Auswahl des in dieser Untersuchung genutzten Modells. GPT-4o-Mini war während eines Großteils der Arbeit das Modell mit dem besten Preis-Leistungs-Verhältnis. Doch bereits während der elf Wochen der Untersuchung kamen leistungsstärkere Modelle wie das chinesische LLM DeepSeek R1 und Googles Gemini 2.0 Flash auf den Markt. Diese neuen Reasoning-Modelle bieten nicht nur eine bessere Performance, sondern sind teilweise auch kostengünstiger. Solche Entwicklungen verdeutlichen, wie schnell sich die technologische Landschaft verändert und wie wichtig es ist, die Wahl der Technologie kontinuierlich zu hinterfragen.

Ein weiterer limitierender Faktor war die begrenzte Verfügbarkeit finanzieller Ressourcen, die es nicht erlaubte, größere Datenmengen zu analysieren. Die Kosten für API-Anfragen und die Verarbeitung umfangreicher Datenmengen waren finanziell nicht tragbar, was dazu führte, dass nur 30 Aktien über einen Zeitraum von drei Monaten untersucht werden konnten. Diese Einschränkung wirkt sich direkt auf die Generalisierbarkeit der Ergebnisse aus. Eine größere Datenbasis, die mehr Unternehmen und einen längeren Zeitraum umfasst, könnte zu robusteren und aussagekräftigeren Ergebnissen führen. Zukünftige Studien könnten von günstigeren API-Anfragen und einer breiteren Datenbasis profitieren, um die Aussagekraft der Analysen zu erhöhen und verschiedene Marktlagen abzudecken.

Die Qualität der verwendeten Daten, insbesondere der Nachrichten, die für die Stimmungsanalyse herangezogen wurden, stellt eine weitere Herausforderung dar. Nachrichtenquellen sind nicht immer neutral, und ihre Qualität kann stark variieren. Dies wirft die Frage auf, ob die Ergebnisse der Analyse mit anderen Nachrichtenquellen ähnlich ausgefallen wären oder ob die Wahl der Quellen einen signifikanten Einfluss auf die Ergebnisse hatte. Zudem zeigte sich, dass kleinere Unternehmen in den Nachrichten oft unterrepräsentiert sind, was die Analyse zusätzlich erschwert. Es wurde beispielsweise festgestellt, dass nur 15 der 30 untersuchten Unternehmen in den ersten zwei Wochen des Jahres 2024 überhaupt relevante Artikel mit einer Relevanzbewertung von über 0,5 auf GNews hatten. Dies verdeutlicht, wie stark die Verfügbarkeit von Daten die Aussagekraft der Ergebnisse beeinflusst. Hinzu kommt, dass beim manuellen Crawlen von Internetseiten, die von Alphavantage zurückgegeben wurden, festgestellt wurde, dass viele Webseiten den Zugriff durch automatisierte Systeme blockieren. Diese Einschränkungen führten dazu, dass einige Nachrichtenquellen nicht präzise von ChatGPT eingestuft werden konnten, da der Inhalt der Artikel fehlte. In solchen Fällen basierte die Einschätzung lediglich auf den Titeln, was als unzureichend bewertet werden muss.

Ein Ansatz, um die Vielfalt der Nachrichtenquellen zu verbessern, könnte die Einbeziehung von Social-Media-Daten, wie beispielsweise Twitter-Nachrichten, sein. Solche nutzergenerierten Inhalte könnten

dazu beitragen, eine größere Meinungsvielfalt und eine differenziertere Stimmungslage in die Analyse einzubeziehen. In der Literatur gibt es bereits Studien [7,40], die Social-Media-Daten erfolgreich genutzt haben und damit bessere Ergebnisse erzielten, auch wenn diese nicht ausschließlich auf der Stimmungsanalyse mithilfe eines LLMs basierten. Ob die Unterschiede in den Ergebnissen jedoch allein auf die Wahl der Nachrichtenquellen zurückzuführen sind, ist unklar. Es ist denkbar, dass auch methodische Probleme wie der sogenannte „Data Snooping“-Effekt eine Rolle spielen. Dieser Effekt beschreibt die Gefahr, dass durch wiederholtes Testen und Anpassen von Modellen an dieselben Daten zufällige Muster als signifikant interpretiert werden. In der Literatur wird dieser Effekt unter anderem von Schwert [66] thematisiert, der darauf hinweist, dass viele Studien dazu neigen, positive Ergebnisse zu bevorzugen, während kritische Perspektiven oft unterrepräsentiert sind.

Im Vergleich zu anderen Studien, wie beispielsweise Lopez-Lira und Tang [55], die eine durchschnittliche Rendite von 0,187 % bei der Stimmungsanalyse berichten, oder Lee et al. [15], die eine Trefferquote von 54 % erzielen, liegen die Ergebnisse dieser Arbeit geringfügig unter den genannten Werten. Es bleibt jedoch unklar, ob in den genannten Studien ein Data-Snooping-Bias vorliegt, der möglicherweise zu den hohen durchschnittlichen Ergebnissen geführt hat. Wie in der vorliegenden Untersuchung gezeigt wird, können sowohl die durchschnittliche Rendite als auch die Trefferquote durch eine gezielte Auswahl von Unternehmen signifikant verbessert werden. Nichtsdestotrotz zeigt sich deutlich, dass die technische Analyse in dieser Arbeit die Stimmungsanalyse in ihrer Rendite deutlich übertrifft.

Die Ergebnisse dieser Arbeit haben auch praktische Implikationen für Investoren. Es zeigte sich, dass der maschinengestützte Handel mit Aktien zwar vielversprechend ist, jedoch mit erheblichen Herausforderungen verbunden bleibt. Besonders für kleinere Unternehmen, die in den Nachrichten weniger präsent sind, ist es unsicher, auf Basis von Stimmungsanalysen fundierte Entscheidungen zu treffen. Für viele Anleger könnte daher der klassische „Buy-and-Hold“-Ansatz weiterhin die einfachere und effektivere Strategie sein, wie auch in der Literatur, beispielsweise von Odean [67] in „Do Investors Trade Too Much?“, diskutiert wird.

Das Zusammenspiel von KI und Finanzmärkten ist ein spannendes, aber auch herausforderndes Forschungsfeld, das noch viele offene Fragen birgt. Zukünftige Studien könnten in den folgenden Bereichen ansetzen, um die Erkenntnisse zu vertiefen und die bestehenden Limitationen zu überwinden.

Erstens könnte eine Erweiterung der Datenbasis und die Durchführung langfristiger Studien dazu beitragen, die Generalisierbarkeit der Ergebnisse zu erhöhen. Mehr Unternehmen müssten über einen längeren Zeitraum und unter verschiedenen Marktlagen untersucht werden, um robustere Aussagen treffen zu können.

Zweitens sollte ein stärkerer Fokus auf kritische Perspektiven gelegt werden, um nicht nur die Stärken, sondern auch die Schwächen und Grenzen der Modelle zu beleuchten.

Drittens wäre es sinnvoll, die neuesten Modelle zu nutzen und deren Ergebnisse miteinander zu vergleichen, um die Auswirkungen technologischer Fortschritte besser zu verstehen.

Schließlich könnte die Einbeziehung alternativer Datenquellen, wie Social Media, die Vielfalt und Aussagekraft der Analysen erhöhen.

Zusammenfassend lässt sich sagen, dass die vorliegende Arbeit wichtige Einblicke in ein dynamisches und hochaktuelles Forschungsfeld liefert. Gleichzeitig wird deutlich, dass noch viele Fragen offen sind

und weitere Forschung notwendig ist, um die Potenziale und Grenzen von KI im Finanzsektor besser zu verstehen und nutzbar zu machen.

8 Literaturverzeichnis

1. Schlude A, Schwind M, Mendel U, Stürz RA, Harles D, Fischer M. Verbreitung und Akzeptanz generativer KI in Deutschland und an deutschen Arbeitsplätzen [Internet]. Bayerisches Forschungsinstitut für Digitale Transformation (bidt); 2023 Nov [zitiert 14. November 2024]. Verfügbar unter: <https://www.bidt.digital/publikation/verbreitung-und-akzeptanz-generativer-ki-in-deutschland-und-an-deutschen-arbeitsplaetzen/>
2. Li Y, Wang S, Ding H, Chen H. Large Language Models in Finance: A Survey. In: 4th ACM International Conference on AI in Finance [Internet]. Brooklyn NY USA: ACM; 2023 [zitiert 14. November 2024]. S. 374–82. Verfügbar unter: <https://dl.acm.org/doi/10.1145/3604237.3626869>
3. Zheng S, Huang J, Chang KCC. Why Does ChatGPT Fall Short in Providing Truthful Answers? [Internet]. arXiv; 2023 [zitiert 14. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2304.10513>
4. Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, u. a. A Survey of Large Language Models [Internet]. arXiv; 2023 [zitiert 14. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2303.18223>
5. Osterrieder J. A Primer on Artificial Intelligence and Machine Learning for the Financial Services Industry. SSRN Electron J [Internet]. 2023 [zitiert 14. November 2024]; Verfügbar unter: <https://www.ssrn.com/abstract=4349078>
6. O'Connor MC. On the Usefulness of Financial Ratios to Investors in Common Stock. Account Rev. 1973;48(2):339–52.
7. Xu Y, Cohen SB. Stock Movement Prediction from Tweets and Historical Prices. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) [Internet]. Melbourne, Australia: Association for Computational Linguistics; 2018 [zitiert 14. November 2024]. S. 1970–9. Verfügbar unter: <http://aclweb.org/anthology/P18-1183>
8. Tam ZR, Wu CK, Tsai YL, Lin CY, Lee H yi, Chen YN. Let Me Speak Freely? A Study on the Impact of Format Restrictions on Performance of Large Language Models [Internet]. arXiv; 2024 [zitiert 21. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2408.02442>
9. Wang Z, Chu Z, Doan TV, Ni S, Yang M, Zhang W. History, development, and principles of large language models: an introductory survey. AI Ethics [Internet]. 14. Oktober 2024 [zitiert 18. November 2024]; Verfügbar unter: <https://link.springer.com/10.1007/s43681-024-00583-7>
10. Goodfellow I, Bengio Y, Courville A. Deep Learning [Internet]. MIT Press; 2016 [zitiert 18. November 2024]. Verfügbar unter: <http://www.deeplearningbook.org>
11. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, u. a. Language Models are Few-Shot Learners [Internet]. arXiv; 2020 [zitiert 14. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2005.14165>
12. Wu S, Irsoy O, Lu S, Dabrowski V, Dredze M, Gehrmann S, u. a. BloombergGPT: A Large Language Model for Finance [Internet]. arXiv; 2023 [zitiert 14. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2303.17564>
13. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving Language Understanding by Generative Pre-Training. 2018 [zitiert 18. November 2024]; Verfügbar unter: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

14. Li X, Chan S, Zhu X, Pei Y, Ma Z, Liu X, u. a. Are ChatGPT and GPT-4 General-Purpose Solvers for Financial Text Analytics? A Study on Several Typical Tasks [Internet]. arXiv; 2023 [zitiert 14. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2305.05862>
15. Lee J, Stevens N, Han SC, Song M. A Survey of Large Language Models in Finance (FinLLMs) [Internet]. arXiv; 2024 [zitiert 14. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2402.02315>
16. Min S, Lyu X, Holtzman A, Artetxe M, Lewis M, Hajishirzi H, u. a. Rethinking the Role of Demonstrations: What Makes In-Context Learning Work? [Internet]. arXiv; 2022 [zitiert 18. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2202.12837>
17. Gallegos IO, Rossi RA, Barrow J, Tanjim MM, Kim S, Dernoncourt F, u. a. Bias and Fairness in Large Language Models: A Survey. *Comput Linguist*. 1. September 2024;50(3):1097–179.
18. Krause D. Large Language Models and Generative AI in Finance: An Analysis of ChatGPT, Bard, and Bing AI. SSRN Electron J [Internet]. 2023 [zitiert 14. November 2024]; Verfügbar unter: <https://www.ssrn.com/abstract=4511540>
19. Kumar D, Jain U, Agarwal S, Harshangi P. Investigating Implicit Bias in Large Language Models: A Large-Scale Study of Over 50 LLMs [Internet]. arXiv; 2024 [zitiert 18. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2410.12864>
20. Eulerich M, Sanatizadeh A, Vakilzadeh H, Wood DA. Can Artificial Intelligence Pass Accounting Certification Exams? ChatGPT: CPA, CMA, CIA, and EA? SSRN Electron J [Internet]. 2023 [zitiert 14. November 2024]; Verfügbar unter: <https://www.ssrn.com/abstract=4452175>
21. Kim A, Muhn M, Nikolaev V. Financial Statement Analysis with Large Language Models [Internet]. arXiv; 2024 [zitiert 14. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2407.17866>
22. Qiu XY, Srinivasan P, Street N. Exploring the Forecasting Potential of Company Annual Reports. *Proc Am Soc Inf Sci Technol*. Januar 2006;43(1):1–15.
23. Mirzadeh I, Alizadeh K, Shahrokhi H, Tuzel O, Bengio S, Farajtabar M. GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models [Internet]. arXiv; 2024 [zitiert 22. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2410.05229>
24. Artificial Analysis [Internet]. [zitiert 12. Februar 2025]. Independent analysis of AI models and API providers. Verfügbar unter: <https://artificialanalysis.ai/?models=o1%2C01-mini%2Cgpt-4o%2Cgpt-4o-mini%2Cllama-3-1-instruct-405b%2Cllama-3-1-instruct-8b%2Cgemini-1-5-pro%2Cgemini-1-5-flash%2Cgemini-1-5-pro-may-2024%2Cclaude-35-sonnet%2Cdeepseek-v2>
25. Bailyn E. firstpagesage. 2025 [zitiert 12. Februar 2025]. Top Generative AI Chatbots by Market Share January 2025. Verfügbar unter: <https://firstpagesage.com/reports/top-generative-ai-chatbots/>
26. Müller N. FAZ. 2024 [zitiert 12. Februar 2025]. ChatGPT erreicht 2,9 Milliarden monatliche Besucher. Verfügbar unter: <https://www.faz.net/pro/digitalwirtschaft/kuenstliche-intelligenz/neuer-rekord-chatgpt-erreicht-2-9-milliarden-besucher-pro-monat-19845849.html>
27. Humlum A, Vestergaard E. The Adoption of ChatGPT. SSRN Electron J [Internet]. 2024 [zitiert 12. Februar 2025]; Verfügbar unter: <https://www.ssrn.com/abstract=4807516>
28. Brynjolfsson E, Li D, Raymond L. Generative AI at Work. *Q J Econ*. 4. Februar 2025;qjae044.
29. OpenAI Platform [Internet]. [zitiert 22. November 2024]. Prompt engineering. Verfügbar unter: <https://platform.openai.com/docs/guides/prompt-engineering>

30. OpenAI Platform [Internet]. [zitiert 27. Dezember 2024]. What are tokens and how to count them? Verfügbar unter: <https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them>
31. Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y. Large Language Models are Zero-Shot Reasoners [Internet]. arXiv; 2022 [zitiert 22. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2205.11916>
32. Krause D. Proper Generative AI Prompting for Financial Analysis. SSRN Electron J [Internet]. 2023 [zitiert 14. November 2024]; Verfügbar unter: <https://www.ssrn.com/abstract=4453664>
33. Wang X, Wei J, Schuurmans D, Le Q, Chi E, Narang S, u. a. Self-Consistency Improves Chain of Thought Reasoning in Language Models [Internet]. arXiv; 2022 [zitiert 22. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2203.11171>
34. Berk J, DeMarzo P, Harford J, Ford G, Mollica V, Finch N. Fundamentals of Corporate Finance. 2. Aufl. Pearson; 2014.
35. Enriques L, Kraakman R. The Anatomy of Corporate Law: A Comparative and Functional Approach. Am J Comp Law. 1. Oktober 2004;52(4):1011.
36. Investor.gov [Internet]. [zitiert 29. November 2024]. Stocks. Verfügbar unter: <https://www.investor.gov/introduction-investing/investing-basics/investment-products/stocks>
37. Securities Act of 1933. US Governement; 2023.
38. Chen J. Essentials of Technical Analysis for Financial Markets. Bd. 1. Wiley; 2010.
39. Brock W, Lakonishok J, LeBARON B. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. J Finance. Dezember 1992;47(5):1731–64.
40. Wu H, Zhang W, Shen W, Wang J. Hybrid Deep Sequential Modeling for Social Text-Driven Stock Prediction. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management [Internet]. Torino Italy: ACM; 2018 [zitiert 14. November 2024]. S. 1627–30. Verfügbar unter: <https://weizhangltt.github.io/paper/cikm18-stock-preprint.pdf>
41. Taylor MP, Allen H. The use of technical analysis in the foreign exchange market. J Int Money Finance. Juni 1992;11(3):304–14.
42. Fama EF. Efficient Capital Markets A Review of Theory and Empirical Work. In: The Fama Portfolio: Selected Papers of Eugene F Fama [Internet]. Chicago: University of Chicago Press; 2017. S. 76–121. Verfügbar unter: <https://www.degruyter.com/document/doi/10.7208/9780226426983-007/pdf?licenseType=restricted#Vancouver>
43. Lo AW. The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective. Forthcoming. 15. Oktober 2004;33.
44. Timmermann A, Granger CWJ. Efficient market hypothesis and forecasting. Int J Forecast. Januar 2004;20(1):15–27.
45. Lewellen J. Predicting returns with financial ratios. J Financ Econ. November 2004;74(2):209–35.
46. Bouwman MJ, Frishkoff PA, Frishkoff P. How do financial analysts make decisions? A process model of the investment screening decision. Account Organ Soc. Januar 1987;12(1):1–29.
47. Engle RF, Ng VK. Measuring and Testing the Impact of News on Volatility. J Finance. Dezember 1993;48(5):1749–78.

48. Petrusheva N, Jordanoski I. Comparative analysis between the fundamental and technical analysis of stocks. *J Oc Process Manag.* 2016;4(2).
49. Sullivan R, Timmermann A, White H. Data-Snooping, Technical Trading Rule Performance, and the Bootstrap. *J Finance.* Oktober 1999;54(5):1647–91.
50. Han Y, Liu Y, Zhou G, Zhu Y. Technical Analysis in the Stock Market: A Review. In: *Handbook of Investment Analysis, Portfolio Management, and Financial Derivatives* [Internet]. WORLD SCIENTIFIC; 2024 [zitiert 18. Dezember 2024]. S. 1893–928. Verfügbar unter: https://www.worldscientific.com/doi/10.1142/9789811269943_0059
51. Welles Wilder J. *New Concepts in Technical Trading Systems.* 1978. 141 S.
52. Grossberg S. Recurrent neural networks. *Scholarpedia.* 2013;8(2):1888.
53. Graves A. Supervised Sequence Labelling with Recurrent Neural Networks [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012 [zitiert 21. November 2024]. (*Studies in Computational Intelligence*; Bd. 385). Verfügbar unter: <https://link.springer.com/10.1007/978-3-642-24797-2>
54. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, u. a. Attention is all you need. Curran Assoc Inc. 4. Dezember 2017;11.
55. Lopez-Lira A, Tang Y. Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models [Internet]. arXiv; 2023 [zitiert 14. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2304.07619>
56. K P, Rudagi S, M N, Patil R, Wadi R. Comparative Study: Stock Prediction Using Fundamental and Technical Analysis. In: *2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)* [Internet]. Tumkur, Karnataka, India: IEEE; 2021 [zitiert 14. November 2024]. S. 1–4. Verfügbar unter: <https://ieeexplore.ieee.org/document/9688449/>
57. Deng Y, He X, Hu J, Yiu SM. Enhancing Few-Shot Stock Trend Prediction with Large Language Models [Internet]. arXiv; 2024 [zitiert 14. November 2024]. Verfügbar unter: <https://arxiv.org/abs/2407.09003>
58. Zhou Z, Ma L, Liu H. Trade the Event: Corporate Events Detection for News-Based Event-Driven Trading. In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* [Internet]. Online: Association for Computational Linguistics; 2021 [zitiert 14. November 2024]. S. 2114–24. Verfügbar unter: <https://aclanthology.org/2021.findings-acl.186>
59. Nakano M, Yamaoka T. Enhancing Sentiment Analysis based Investment by Large Language Models in Japanese Stock Market. *SSRN Electron J* [Internet]. 2023 [zitiert 14. November 2024]; Verfügbar unter: <https://www.ssrn.com/abstract=4511658>
60. Wisniewski TP, Yekini LS. Stock market returns and the content of annual report narratives. *Account Forum.* Dezember 2015;39(4):281–94.
61. Sun L. The Motley Fool. 2023 [zitiert 3. Januar 2025]. 3 Things About Arm Holdings That Smart Investors Know. Verfügbar unter: <https://www.fool.com/investing/2023/10/05/things-about-arm-holdings-smart-investors-know/>
62. OpenAI Platform [Internet]. [zitiert 3. Januar 2024]. Rate Limits. Verfügbar unter: <https://platform.openai.com/docs/guides/rate-limits?context=tier-one>
63. Metrick A. Performance Evaluation with Transactions Data: The Stock Selection of Investment Newsletters. *J Finance.* Oktober 1999;54(5):1743–75.

64. Grinblatt M, Moskowitz TJ. Predicting stock price movements from past returns: the role of consistency and tax-loss selling. *J Financ Econ.* März 2004;71(3):541–79.
65. Kenton W. investopedia. 2021 [zitiert 15. Februar 2025]. Profit/Loss Ratio Definition, Formula, How It Works. Verfügbar unter:
https://www.investopedia.com/terms/p/profit_loss_ratio.asp#:~:text=The%20profit%2Floss%20ratio%20is,over%20a%20specified%20time%20period.
66. Schwert G. Anomalies and Market Efficiency. *Handb Econ Finance.* 2003;1.
67. Odean T. Do Investors Trade Too Much? *Am Econ Rev.* 1. Dezember 1999;89(5):1279–98.

9 Anhang

```

1   "duration": "0.7131867924000005 seconds",
2   "data": [
3     {
4       "title": "Alphabet Options: A Look at what the Big Money is Thinking - Alphabet ( NASDAQ:GOOGL )",
5       "url": "https://www.benzinga.com/insights/options/24/12/42726409/alphabets-options-a-look-at-what-the-big-money-is-thinking",
6       "summary": "Investors with a lot of money to spend have taken a bullish stance on Alphabet (GOOGL). We noticed this today when the trades showed up",
7       "sentimentScore": "0.00097",
8       "relevanceScore": "-0.38985",
9     },
10    {
11      "title": "Alphabet CEO Pichai Warns Employees Of High Stakes 2025: 'We Are Facing Scrutiny Across The World' - Alphabet ( NASDAQ:GOOGL )",
12      "url": "https://www.benzinga.com/reach/24/12/42728738/alphabet-ceo-pichai-warns-employees-of-high-stakes-2025-we-are-facing-scrutiny-across-the-world",
13      "summary": "CEO Sundar Pichai shared warnings for 2025 to the company's more than 100,000 employees recently. Regulatory pressure and t",
14      "sentimentScore": "0.223855",
15      "relevanceScore": "-0.38985",
16    },
17    {
18      "title": "Watch the Official Delta Force 'Serpentine' PkD Raid Trailer",
19      "url": "https://www.benzinga.com/pressreleases/24/12/42726409/watch-the-official-delta-force-serpentine-pkd-raid-trailer",
20      "summary": "LOS ANGELES, Dec. 30, 2024 //Newswire/ -- Prepare for an all-new type of Delta Force challenge, Operation Serpentine drops you and",
21      "sentimentScore": "0.241645",
22      "relevanceScore": "-0.38985",
23    },
24    {
25      "title": "Trump Administration to Challenge Social Media Censorship, Setting Stage for Conflict with EU Regulations - Alphabet ( NASDAQ:GOOGL )",
26      "url": "https://www.benzinga.com/24/12/42729529/social-media-stocks-react-as-trump-targets-censorship-and-delays-tiktok-ban",
27      "summary": "The Trump administration's 'censorship cartel,' urging FCC and FTC to restrict social media content moderation, Europe enforces stricter diploma",
28      "sentimentScore": "0.299717",
29      "relevanceScore": "-0.261117",
30    },
31    {
32      "title": "Comment Stock Rises 26% in Six Months; What Should Investors Do?",
33      "url": "https://www.zacks.com/stocks/reviews/339914/comment-stock-rises-26-in-six-months-what-should-investors-do",
34      "summary": "COLD stock offers compelling long-term growth opportunities, yet its elevated valuation suggests that short-term gains may be constr",
35      "sentimentScore": "0.331169",
36      "relevanceScore": "-0.212594",
37    },
38    {
39      "title": "The Stock Picks For Week of December 30, 2024",
40      "url": "https://www.zacks.com/stock/news/339652/the-stock-picks-for-week-of-december-30-2024",
41      "summary": "A Cash-Flow Company That's Become a Shining Star and a Mobile App Company That's Outperformed Its Sector So Far This Year.",
42      "sentimentScore": "0.318165",
43      "relevanceScore": "-0.119895",
44    },
45    {
46      "title": "Rezerve AI Converts $59 Million in Variable Rate Convertible Loans Before Year-End, Strengthening Financial Position for 2025 - Rezerve AI Converts $59 Million in Variable Rate Convertible Loans Before Year-End Summary - NEW YORK, Dec. 30, 2024 | GLOBE NEWSWIRE | — Rezerve AI (REZV), a leader in AI-driven retail solutions, announces the successful conv",
47      "sentimentScore": "0.262715",
48      "relevanceScore": "-0.168958",
49    },
50    {
51      "title": "Square Researchers Expose OAuth Attack on Chrome Extensions Days Before Major Breach",
52      "url": "https://www.benzinga.com/pressreleases/24/12/42729457/square-researchers-expose-oauth-attack-on-chrome-extensions-days-before-major-br",
53      "summary": "PALO ALTO, Calif., Dec. 30, 2024 ( GLOBE NEWSWIRE ) — Square, an industry-first Brewer Detection and Response ( BDR ) solution, L",
54      "sentimentScore": "-0.132279",
55      "relevanceScore": "-0.000000",
56    },
57  ],
58  {
59    "reasons_array": [
60      {
61        "id": "16046",
62        "reasoning": "The article discusses Apple's discontinuation of a popular accessory, which may negatively impact accessory sales and investor sentiment, reflected in a sentiment score of -0.15."
63      },
64      {
65        "id": "13496",
66        "reasoning": "The price cut on the Apple Watch 10 is likely to boost sales and consumer interest, contributing positively to investor sentiment with a sentiment score of 0.8."
67      },
68      {
69        "id": "13502",
70        "reasoning": "Highlighting the anniversary of a safety feature on iPhones, this article positively reinforces Apple's commitment to user safety, with a sentiment score of 0.6."
71      },
72      {
73        "id": "13503",
74        "reasoning": "The addition of older products to the vintage list indicates product lifecycle management but does not significantly impact current sales, with a sentiment score of -0.2."
75      },
76      {
77        "id": "13506",
78        "reasoning": "The introduction of new features in tvOS 18 enhances user experience, likely leading to increased customer satisfaction and loyalty, with a sentiment score of 0.6."
79      },
80      {
81        "id": "16047",
82        "reasoning": "The announcement of AirTag 2 indicates innovation and potential growth in the accessories market, contributing positively to investor sentiment with a sentiment score of 0.6."
83      },
84      {
85        "id": "13495",
86        "reasoning": "The discussion of Apple's content offerings through Apple TV+ is relevant but does not significantly impact stock performance, with a sentiment score of 0.3."
87      },
88      {
89        "id": "13501",
90        "reasoning": "The high review score for the M4 MacBook Pro indicates strong product performance and positive consumer reception, with a sentiment score of 0.85."
91      },
92      {
93        "id": "13499",
94        "reasoning": "The improvements in iOS 18's Messages app are positive but mention of a needed fix may temper enthusiasm, with a sentiment score of 0.3."
95      },
96      {
97        "id": "13497",
98        "reasoning": "The significant discount on AirPods during Black Friday is likely to drive sales, positively impacting investor sentiment with a sentiment score of 0.75."
99      }
100    ],
101    "confidence_score": 0.65,
102    "reasoning_summary": "The overall sentiment from recent news articles about AAPL is mixed, with several positive developments such as product launches and discounts, but also some negative news regarding product discontinuations and market performance. The positive sentiment from product innovations and sales promotions suggests a potential for growth, while negative sentiments from accessory sales concerns and market competition temper the outlook. Overall, the sentiment leans slightly positive, indicating a LONG position may be favorable.",
103    "action": "long"
104  }

```

Abbildung 15A: Vergleich von den Stimmungs- und Relevanzscores von Alphavantage (links) und ChatGPT (rechts)

```

{
  "reasons_array": [
    {
      "id": "16046",
      "reasoning": "The article discusses Apple's discontinuation of a popular accessory, which may negatively impact accessory sales and investor sentiment, reflected in a sentiment score of -0.15."
    },
    {
      "id": "13496",
      "reasoning": "The price cut on the Apple Watch 10 is likely to boost sales and consumer interest, contributing positively to investor sentiment with a sentiment score of 0.8."
    },
    {
      "id": "13502",
      "reasoning": "Highlighting the anniversary of a safety feature on iPhones, this article positively reinforces Apple's commitment to user safety, with a sentiment score of 0.6."
    },
    {
      "id": "13503",
      "reasoning": "The addition of older products to the vintage list indicates product lifecycle management but does not significantly impact current sales, with a sentiment score of -0.2."
    },
    {
      "id": "13506",
      "reasoning": "The introduction of new features in tvOS 18 enhances user experience, likely leading to increased customer satisfaction and loyalty, with a sentiment score of 0.6."
    },
    {
      "id": "16047",
      "reasoning": "The announcement of AirTag 2 indicates innovation and potential growth in the accessories market, contributing positively to investor sentiment with a sentiment score of 0.6."
    },
    {
      "id": "13495",
      "reasoning": "The discussion of Apple's content offerings through Apple TV+ is relevant but does not significantly impact stock performance, with a sentiment score of 0.3."
    },
    {
      "id": "13501",
      "reasoning": "The high review score for the M4 MacBook Pro indicates strong product performance and positive consumer reception, with a sentiment score of 0.85."
    },
    {
      "id": "13499",
      "reasoning": "The improvements in iOS 18's Messages app are positive but mention of a needed fix may temper enthusiasm, with a sentiment score of 0.3."
    },
    {
      "id": "13497",
      "reasoning": "The significant discount on AirPods during Black Friday is likely to drive sales, positively impacting investor sentiment with a sentiment score of 0.75."
    }
  ],
  "confidence_score": 0.65,
  "reasoning_summary": "The overall sentiment from recent news articles about AAPL is mixed, with several positive developments such as product launches and discounts, but also some negative news regarding product discontinuations and market performance. The positive sentiment from product innovations and sales promotions suggests a potential for growth, while negative sentiments from accessory sales concerns and market competition temper the outlook. Overall, the sentiment leans slightly positive, indicating a LONG position may be favorable.",
  "action": "long"
}

```

Abbildung 16A: Analyseergebnis von ChatGPT mit Handelsempfehlung und den Begründungen für die Entscheidung

```

Company: AAPL
Analysis Type: technical
Overall Wins:
| Max: 24, Min: 23, Average: 23.20, Std Dev: 0.40
Overall Losses:
| Max: 2, Min: 1, Average: 1.80, Std Dev: 0.40
Closing Strategy Breakdown:
Strategy: day_3_close
| Wins: Max: 5, Min: 5, Average: 5.00, Std Dev: 0.00
| Losses: Max: 0, Min: 0, Average: 0.00, Std Dev: 0.00
Strategy: day_2_close
| Wins: Max: 5, Min: 5, Average: 5.00, Std Dev: 0.00
| Losses: Max: 0, Min: 0, Average: 0.00, Std Dev: 0.00
Strategy: day_1_close
| Wins: Max: 5, Min: 5, Average: 5.00, Std Dev: 0.00
| Losses: Max: 0, Min: 0, Average: 0.00, Std Dev: 0.00
Strategy: day_close
| Wins: Max: 5, Min: 5, Average: 5.00, Std Dev: 0.00
| Losses: Max: 0, Min: 0, Average: 0.00, Std Dev: 0.00
Strategy: take_profit
| Wins: Max: 4, Min: 3, Average: 3.20, Std Dev: 0.40
| Losses: Max: 2, Min: 1, Average: 1.80, Std Dev: 0.40

Analysis Type: sentiment
Overall Wins:
| Max: 16, Min: 12, Average: 13.60, Std Dev: 1.96
Overall Losses:
| Max: 4, Min: 4, Average: 4.00, Std Dev: 0.00
Closing Strategy Breakdown:
Strategy: day_3_close
| Wins: Max: 4, Min: 3, Average: 3.40, Std Dev: 0.49
| Losses: Max: 1, Min: 1, Average: 1.00, Std Dev: 0.00
Strategy: day_2_close
| Wins: Max: 4, Min: 3, Average: 3.40, Std Dev: 0.49
| Losses: Max: 1, Min: 1, Average: 1.00, Std Dev: 0.00
Strategy: day_1_close
| Wins: Max: 4, Min: 3, Average: 3.40, Std Dev: 0.49
| Losses: Max: 1, Min: 1, Average: 1.00, Std Dev: 0.00
Strategy: day_close
| Wins: Max: 4, Min: 3, Average: 3.40, Std Dev: 0.49
| Losses: Max: 1, Min: 1, Average: 1.00, Std Dev: 0.00

```

Abbildung 17A: Standardabweichung AAPL

```
Company: XOM
  Analysis Type: technical
    Overall Wins:
      Max: 11, Min: 10, Average: 10.20, Std Dev: 0.40
    Overall Losses:
      Max: 15, Min: 14, Average: 14.80, Std Dev: 0.40
  Closing Strategy Breakdown:
    Strategy: day_3_close
      Wins: Max: 1, Min: 1, Average: 1.00, Std Dev: 0.00
      Losses: Max: 4, Min: 4, Average: 4.00, Std Dev: 0.00
    Strategy: day_2_close
      Wins: Max: 3, Min: 3, Average: 3.00, Std Dev: 0.00
      Losses: Max: 2, Min: 2, Average: 2.00, Std Dev: 0.00
    Strategy: day_1_close
      Wins: Max: 3, Min: 3, Average: 3.00, Std Dev: 0.00
      Losses: Max: 2, Min: 2, Average: 2.00, Std Dev: 0.00
    Strategy: day_close
      Wins: Max: 3, Min: 3, Average: 3.00, Std Dev: 0.00
      Losses: Max: 2, Min: 2, Average: 2.00, Std Dev: 0.00
    Strategy: take_profit
      Wins: Max: 1, Min: 0, Average: 0.20, Std Dev: 0.40
      Losses: Max: 5, Min: 4, Average: 4.80, Std Dev: 0.40
```

Abbildung 18A: Standardabweichung XOM

```

Company: DKNG
  Analysis Type: technical
    Overall Wins:
      Max: 18, Min: 12, Average: 15.60, Std Dev: 2.24
    Overall Losses:
      Max: 13, Min: 7, Average: 9.40, Std Dev: 2.24
    Closing Strategy Breakdown:
      Strategy: day_3_close
        Wins: Max: 4, Min: 2, Average: 3.20, Std Dev: 0.75
        Losses: Max: 3, Min: 1, Average: 1.80, Std Dev: 0.75
      Strategy: day_2_close
        Wins: Max: 4, Min: 2, Average: 3.20, Std Dev: 0.75
        Losses: Max: 3, Min: 1, Average: 1.80, Std Dev: 0.75
      Strategy: day_1_close
        Wins: Max: 4, Min: 2, Average: 3.20, Std Dev: 0.75
        Losses: Max: 3, Min: 1, Average: 1.80, Std Dev: 0.75
      Strategy: day_close
        Wins: Max: 4, Min: 2, Average: 2.80, Std Dev: 0.75
        Losses: Max: 3, Min: 1, Average: 2.20, Std Dev: 0.75
      Strategy: take_profit
        Wins: Max: 4, Min: 2, Average: 3.20, Std Dev: 0.75
        Losses: Max: 3, Min: 1, Average: 1.80, Std Dev: 0.75

  Analysis Type: sentiment
    Overall Wins:
      Max: 3, Min: 3, Average: 3.00, Std Dev: 0.00
    Overall Losses:
      Max: 1, Min: 1, Average: 1.00, Std Dev: 0.00
    Closing Strategy Breakdown:
      Strategy: day_3_close
        Wins: Max: 1, Min: 1, Average: 1.00, Std Dev: 0.00
        Losses: Max: 0, Min: 0, Average: 0.00, Std Dev: 0.00
      Strategy: day_2_close
        Wins: Max: 1, Min: 1, Average: 1.00, Std Dev: 0.00
        Losses: Max: 0, Min: 0, Average: 0.00, Std Dev: 0.00
      Strategy: day_1_close
        Wins: Max: 1, Min: 1, Average: 1.00, Std Dev: 0.00
        Losses: Max: 0, Min: 0, Average: 0.00, Std Dev: 0.00
      Strategy: day_close
        Wins: Max: 0, Min: 0, Average: 0.00, Std Dev: 0.00
        Losses: Max: 1, Min: 1, Average: 1.00, Std Dev: 0.00

```

Abbildung 194: Standardabweichung DKNG

```

Company: IBIO
Analysis Type: technical
Overall Wins:
| Max: 9, Min: 6, Average: 7.40, Std Dev: 1.36
Overall Losses:
| Max: 19, Min: 11, Average: 16.60, Std Dev: 3.01
Closing Strategy Breakdown:
Strategy: day_3_close
| Wins: Max: 3, Min: 0, Average: 0.80, Std Dev: 1.17
| Losses: Max: 5, Min: 2, Average: 4.00, Std Dev: 1.26
Strategy: day_2_close
| Wins: Max: 2, Min: 1, Average: 1.40, Std Dev: 0.49
| Losses: Max: 4, Min: 2, Average: 3.40, Std Dev: 0.80
Strategy: day_1_close
| Wins: Max: 2, Min: 1, Average: 1.80, Std Dev: 0.40
| Losses: Max: 4, Min: 2, Average: 3.00, Std Dev: 0.63
Strategy: day_close
| Wins: Max: 3, Min: 2, Average: 2.80, Std Dev: 0.40
| Losses: Max: 3, Min: 1, Average: 2.00, Std Dev: 0.63
Strategy: take_profit
| Wins: Max: 1, Min: 0, Average: 0.60, Std Dev: 0.49
| Losses: Max: 5, Min: 3, Average: 4.20, Std Dev: 0.75

```

Abbildung 20A: Standardabweichung IBIO

```

Company: MRK
Analysis Type: technical
Overall Wins:
| Max: 16, Min: 8, Average: 12.40, Std Dev: 2.87
Overall Losses:
| Max: 14, Min: 9, Average: 11.60, Std Dev: 1.85
Closing Strategy Breakdown:
Strategy: day_3_close
| Wins: Max: 3, Min: 1, Average: 2.40, Std Dev: 0.80
| Losses: Max: 3, Min: 2, Average: 2.40, Std Dev: 0.49
Strategy: day_2_close
| Wins: Max: 2, Min: 1, Average: 1.60, Std Dev: 0.49
| Losses: Max: 4, Min: 3, Average: 3.20, Std Dev: 0.40
Strategy: day_1_close
| Wins: Max: 5, Min: 3, Average: 4.00, Std Dev: 0.89
| Losses: Max: 2, Min: 0, Average: 0.80, Std Dev: 0.75
Strategy: day_close
| Wins: Max: 5, Min: 3, Average: 4.00, Std Dev: 0.89
| Losses: Max: 2, Min: 0, Average: 0.80, Std Dev: 0.75
Strategy: take_profit
| Wins: Max: 1, Min: 0, Average: 0.40, Std Dev: 0.49
| Losses: Max: 5, Min: 4, Average: 4.40, Std Dev: 0.49

```

Abbildung 21A: Standardabweichung MRK

```

1   "duration": "0.7132787920000002 seconds",
2   "data": [
3     {
4       "title": "Alphabet's Options: A Look at What the Big Money Is Thinking - Alphabet [ NASDAQ:GOOGL ]",
5       "url": "https://www.benzinga.com/insights/options/24/12/42726498/alphabets-options-a-look-at-what-the-big-money-is-thinking",
6       "summary": "Investors with a lot of money to spend have taken a bullish stance on Alphabet (GOOGL). We noticed this today when the trades showed up",
7       "sentimentScore": "0.00007",
8       "relevanceScore": "0.00005",
9     },
10    {
11      "title": "Alphabet CEO Pichai Warns Employees Of High Stakes 2025: 'We Are Facing Scrutiny Across The World' - Alphabet [ NASDAQ:GOOGL ] , Alpha",
12      "url": "https://www.benzinga.com/tech/24/12/42728739/alpha-ceo-pichai-warns-employees-of-high-stakes-2025-we-are-facing-scrutiny-across-the-world",
13      "summary": "Alphabet CEO Sundar Pichai shared warnings for 2025 to the company's more than 100,000 employees recently. Regulatory pressure and t",
14      "sentimentScore": "0.22035",
15      "relevanceScore": "0.30006",
16    },
17    {
18      "title": "Watch the Official Delta Force 'Serpentine' Pic Had Trailer",
19      "url": "https://www.benzinga.com/reviews/24/12/42726765/watch-the-official-delta-force-serpentine-pic-had-trailer",
20      "summary": "LOS ANGELES, Dec. 30, 2024 //PRNewswire/ -- Prepare for an all-new type of Delta Force challenge. Operative Serpentine drops you and",
21      "sentimentScore": "0.01245",
22      "relevanceScore": "0.00007",
23    },
24    {
25      "title": "Trump Administration to Challenge Social Media Censorship, Setting Stage for Conflict with EU Regulations - Alphabet [ NASDAQ:GOOGL ]",
26      "url": "https://www.benzinga.com/tech/24/12/42729529/social-media-stocks-react-as-trump-targets-censorship-and-delays-tiktok-ban",
27      "summary": "The Trump administration is taking a stand against social media platforms like TikTok and Twitter. The US government has filed a complaint against the Chinese-owned app TikTok, alleging it violates US national security laws. The administration is also pushing back against European Union regulations that would ban TikTok from the US market. This could lead to a significant conflict between the US and EU over digital freedom and regulation. The stock market has reacted positively to this news, with Alphabet (GOOGL) and other tech stocks seeing gains. The overall sentiment is mixed, with some investors看好 the potential for a regulatory battle that could benefit the company, while others are concerned about the long-term impact on its business. The relevance score is relatively low, indicating that this story is less newsworthy than others on the list.",",
28      "sentimentScore": "-0.29970",
29      "relevanceScore": "0.26011",
30    },
31    {
32      "title": "Comcast Stock Rises 20% In Six Months; What Should Investors Do",
33      "url": "https://www.zacks.com/stocks/news/33954/comcast-stock-rises-20-in-six-months-what-should-investors-do",
34      "summary": "CIVI stock offers compelling long-term growth opportunities, yet its elevated valuation suggests that short-term gains may be constrain",
35      "sentimentScore": "0.31020",
36      "relevanceScore": "0.31084",
37    },
38    {
39      "title": "Stock Picks for Week of December 30, 2024",
40      "url": "https://www.zacks.com/stock/news/339052/top-stock-picks-for-week-of-december-30-2024",
41      "summary": "A Cash-rich Company That's Become a Shining Star and a Mobile App Company That's Outperformed Its Sector So Far This Year.",",
42      "sentimentScore": "0.18100",
43      "relevanceScore": "0.11983",
44    },
45    {
46      "title": "Revolve AI Converts $59 Million in Variable Rate Convertible Loans Before Year-End, Strengthening Financial Position for 2025 - Revolv",
47      "url": "https://www.benzinga.com/releases/24/12/42722593/revolve-ai-converts-59-million-in-variable-rate-convertible-loans-before-year-end",
48      "summary": "NEW YORK, Dec. 30, 2024 / GLOBE NEWSWIRE / — Revolve AI (RAV), a leader in AI-driven retail solutions, announces the successful conve",
49      "sentimentScore": "0.20271",
50      "relevanceScore": "0.10095",
51    },
52    {
53      "title": "Square Researchers Expose Bluto Attack on Chrome Extensions Days Before Major Break",
54      "url": "https://www.benzinga.com/releases/24/12/42723457/square-researchers-expose-bluto-attack-on-chrome-extensions-days-before-major-br",
55      "summary": "PALO ALTO, Calif., Dec. 30, 2024 / GLOBE NEWSWIRE / — SquareX, an industry-first Browser Detection and Response (BDR) solution, h",
56      "sentimentScore": "-0.13227",
57      "relevanceScore": "0.00007",
58    }
59  ]
60 
```

Abbildung 22A: Vergleich von Stimmungsanalyse Alphavantage (links) und ChatGPT (rechts) bei zu vielen Input Token