



Article

# AI-Driven Financial Analysis: Exploring ChatGPT's Capabilities and Challenges

Li Xian Liu <sup>1,\*</sup>, Zhiyue Sun <sup>2</sup>, Kunpeng Xu <sup>3</sup> and Chao Chen <sup>4</sup>

<sup>1</sup> College of Business, Law & Governance, James Cook University, 1 James Cook Drive, Douglas, QLD 4811, Australia

<sup>2</sup> School of Accounting, Economics & Finance, Curtin University, Kent Street, Perth, WA 6102, Australia; zhiyue.sun@curtin.edu.au

<sup>3</sup> School of Statistics and Information, Shanghai University of International Business and Economics, 1900 Wenxiang Rd, Songjiang District, Shanghai 201613, China; xukunpeng24@163.com

<sup>4</sup> Accounting, Information System and Supply Chain, RMIT University, 124 La Trobe St, Melbourne, VIC 3000, Australia; chao.chen@rmit.edu.au

\* Correspondence: li.liu1@jcu.edu.au

**Abstract:** The transformative impact of AI technologies on the financial sector has been a topic of increasing interest. This study investigates ChatGPT's applications in financial reasoning and analysis and evaluates ChatGPT-4o's effectiveness and limitations in conducting both basic and complex financial analysis tasks. By designing a series of multi-step, advanced reasoning tasks and establishing task-specific evaluation metrics, we assessed ChatGPT-4o's performance compared to human analysts. Results indicate that while ChatGPT-4o demonstrates proficiency in basic and some complex financial tasks, it struggles with deep analytical and critical thinking tasks, especially in specialized finance areas. This study underscores the need for meticulous task formulation and robust evaluation in AI financial applications. While ChatGPT enhances efficiency, integrating it with human expertise is crucial for effective decision-making. Our findings highlight both the potential and limitations of ChatGPT-4o in financial analysis, providing valuable insights for future AI integration in the finance sector.

**Keywords:** financial analysis and reasoning; financial modeling; evaluation metrics; ChatGPT-4o; Human-AI collaboration; AI practical applications in finance

**Citation:** Liu, Li Xian, Zhiyue Sun, Kunpeng Xu, and Chao Chen. 2024. AI-Driven Financial Analysis: Exploring ChatGPT's Capabilities and Challenges. *International Journal of Financial Studies* 12: 60. <https://doi.org/10.3390/ijfs12030060>

Academic Editor: Panagiotis Dimitropoulos

Received: 20 May 2024

Revised: 18 June 2024

Accepted: 25 June 2024

Published: 27 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In 1950, Alan Turing published his seminal paper, "Computing Machinery and Intelligence" (Turing 2012), posing a profound question, "Can machines think?" Over seventy years later, on 30 November 2022, OpenAI launched ChatGPT (Chat Generative Pre-training Transformer), a revolutionary Artificial Intelligence (AI) language model that has rapidly transformed various sectors in a remarkably short time span.

Trained on extensive datasets using advanced Natural Language Processing (NLP) techniques and enhanced by Reinforcement Learning from Human Feedback, ChatGPT can perform a wide array of tasks. Unlike traditional search engines, it provides specific, concise answers and features an advanced data analysis tool. Recent updates have introduced audio and video interaction capabilities, further expanding its functionality. This enables it to write and execute code, perform complex financial analyses, and produce downloadable outputs, making it invaluable for precision and efficiency in financial analysis.

Throughout history, transformative technologies like manufacturing automation and the rise of e-commerce have ushered in new epochs. ChatGPT's rapid adoption reflects this historical pattern. For instance, in Europe, the travel company Expedia has

harnessed AI chatbots to help users plan cost-effective, eco-friendly trips (Blesiada 2023). According to Enterprise Apps Today, the technology and education sectors are among the foremost adopters of OpenAI's solutions, with industries such as business services, manufacturing, and finance also integrating AI into their operations (Elad 2024). A 2023 Goldman Sachs report highlighted AI's potential to displace up to 300 million full-time jobs (Kelly 2023), sparking debates among financial analysts about the future relevance of their roles in an increasingly automated economy.

Despite these benefits, current AI models, including ChatGPT, present some dilemmas. For instance, the accuracy and quality of ChatGPT's responses can vary based on the question posed, the training data available, the complexity of the topic, and the given instructions or prompts (Kocoń et al. 2023). Further, the current AI models, including ChatGPT, still struggle with tasks requiring deep understanding and critical thinking (Roumeliotis and Tselikas 2023). Therefore, evaluating ChatGPT-4o's performance in financial analysis is crucial. Automating financial tasks can enhance efficiency, reduce costs, and provide consistent, objective analysis. Understanding its capabilities and limitations helps address regulatory and ethical concerns, informs workforce transition strategies, and drives innovation. This study aims to investigate ChatGPT-4o's effectiveness in performing financial analysis tasks traditionally handled by human analysts, offering insights into its potential and constraints in the financial sector.

To achieve this objective, we designed a set of multi-step and advanced reasoning financial tasks and established specific evaluation metrics. We then conducted empirical experiments to assess ChatGPT-4o's performance on these tasks compared to human analysts. Our results indicate that while ChatGPT-4o can effectively perform basic and some complex financial tasks, it has limitations in tasks that involve managing complex financial information and specialized finance areas such as derivatives.

This research contributes to the understanding of AI's role in finance by providing insights into ChatGPT's financial applications and highlighting its potential limitations. These findings enhance the knowledge base for academicians, developers, and stakeholders interested in integrating ChatGPT into financial practices in our business world.

Our paper is organized as follows: Section 2 reviews AI-related financial studies, Section 3 outlines the empirical design and tests, Section 4 presents and analyses the findings and discusses their practical applications and implementations, and Section 5 concludes.

## **2. Artificial Intelligence Techniques and Related Studies in Financial Analysis**

### *2.1. Historical Evolution and Technological Advancements*

Technological advancements have profoundly influenced the evolution of the financial services industry. Innovations such as telegrams and Morse code in the late 19th century revolutionized monetary transactions, setting the stage for further technological progress (Saunders et al. 2021). The transition to digital banking in the 20th and 21st centuries marked a pivotal shift, with financial technology (FinTech) fundamentally transforming trading practices and financial management. The 1970s introduced algorithmic trading in financial institutions, leveraging computer models to automate trading strategies. This technological evolution enabled the development of advanced trading models that could analyze extensive datasets, identify patterns, and make informed trading decisions (Burgess 2021).

### *2.2. AI Applications in Financial Analysis*

AI stands out as a significant and expanding field of interest among scholars and practitioners. Its applications extend across traditional areas like financial markets, trading, banking, investments, optimization, and insurance. Additionally, AI is increasingly pivotal in burgeoning FinTech sectors, including big data analytics, blockchain, and data mining. These applications are crucial for risk management and regulatory compliance (Ahmed et al. 2022; Cao 2022; Farooq and Chawla 2021; Lin 2019).

### 2.2.1. Enhancing Market Efficiency and Risk Management

AI-driven trading strategies have shown to outperform human traders under various market conditions including during crises such as the COVID-19 pandemic (Burgess 2021). The integration of machine learning and AI techniques has further refined algorithmic trading, significantly influencing market dynamics, liquidity, and trading strategies, thereby enhancing overall market efficiency (Chaboud et al. 2014). Moreover, advanced algorithms and machine learning models have demonstrated their efficacy in analyzing extensive datasets to identify potential risks (Demajo et al. 2020; Yu et al. 2023) and detect patterns of fraudulent activities (Jullum et al. 2020).

### 2.2.2. Predictive Analytics and Financial Stability

Since the 1990s, AI methodologies such as artificial neural networks, support vector machines, ensemble methods, generalized boosting, AdaBoost, and Random Forests have been employed to predict financial distress and failures in banks (Liu et al. 2021). The implementation of Explainable AI (XAI) in credit models within the banking sector, such as credit scoring and credit default prediction, has facilitated greater transparency and understanding of complex financial concepts, promoting their adoption in the finance industry (Demajo et al. 2020; de Lange et al. 2022).

### 2.2.3. Modeling Behavioural Biases and Sentiment Analysis

The use of AI to model behavioral biases has also gained prominence. The integration of Natural Language Processing (NLP) has become increasingly vital in finance studies since the early 21st century, covering areas such as text classification, sentiment analysis, and natural language generation. Research by Tetlock et al. (2008) and Bollen et al. (2011) has shown the predictive power of sentiment analysis in determining stock market trends, establishing a significant correlation between news sentiment and market behavior. Similarly, Félix et al. (2020) have employed machine learning-based models to construct implied volatility sentiment, further highlighting the utility of AI in financial analytics.

## 2.3. Emergence of ChatGPT in Financial Analysis

Since its inception in November 2022, ChatGPT has sparked considerable academic interest in its application to finance. Researchers have explored its utility in a variety of financial tasks, including financial document classification, sentiment analysis, named entity recognition in financial texts, and financial data extraction (Zaremba and Demir 2023). Traditional keyword-based methods in financial sentiment analysis have shown weaknesses, particularly in handling complex texts, as these methods are susceptible to adversarial manipulation (Boukes et al. 2020; Hartmann et al. 2023; Leippold 2023a).

ChatGPT's ability to interface with explainable AI models and demystify complex financial concepts for lay audiences underscores its potential in enhancing financial analysis and research (Wenzlaff and Spaeth 2022; Yue et al. 2023). However, Leippold (2023b) cautioned that large language models (LLMs) like GPT-3 might generate unfounded content, as demonstrated in tests involving GPT-3's responses on climate change topics. Furthermore, Lopez-Lira and Tang (2023) discovered a significant correlation between ChatGPT's interpretations of corporate news and subsequent stock market reactions, suggesting its accuracy in financial analysis.

In finance research, Dowling and Lucey (2023) highlighted ChatGPT's contributions across various stages of research, particularly in the study of cryptocurrencies. Hansen and Kazinnik (2023) demonstrated ChatGPT's effectiveness in analyzing central bank communications, underscoring its value in comparative studies and zero-shot learning capabilities.

The market for AI in finance is experiencing significant growth and is driven by key players who are facilitating this transformation. Services such as KAI, AlphaChat, Growthbotics, and FinChat have been developed to meet the specific requirements of the

financial sector. FinChat, in particular, leverages generative AI to provide investment research, offering fundamental investors relevant data through an interactive conversational interface.

### Ethical and Regulatory Considerations

Despite its advantages, the deployment of AI models such as ChatGPT in financial settings presents significant ethical and regulatory challenges. Ensuring the responsible use of AI is crucial, particularly in areas of risk management and regulatory compliance (Zaremba and Demir 2023). The increasing acknowledgment of ChatGPT's potential to influence financial practices and research necessitates robust measures to address these challenges and fully harness AI's potential to enhance financial analysis.

## 3. Empirical Design

### 3.1. Financial Analysis and Reasoning in a Nutshell

Financial analysis can range from simple to complex, depending on the context and specific goals of the analysis. It involves the systematic examination of financial data to assess the performance of a business or investment and to make forecasts.

Unlike basic mathematical calculations, most simple financial analyses involve multi-step processes. A prime example is the concept of present value, a fundamental principle in finance widely used to determine the value of shares, bonds, projects, or entire businesses. Calculating present value requires several steps: identifying future cash flows, selecting the appropriate discount rate, determining the number of periods for each cash flow, and computing the present value for each cash flow. Other simple financial analyses include ratio and metric calculations, as well as simple budgeting and forecasting using historical data.

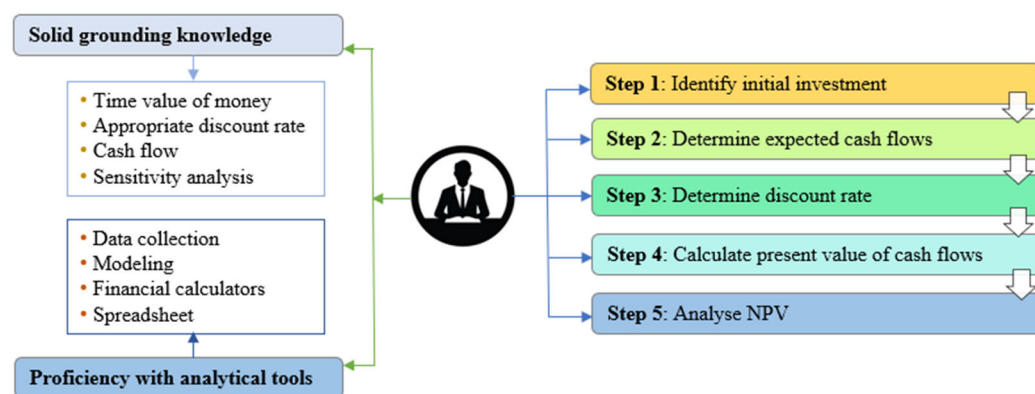
As we progress to more advanced or complex financial analysis, the necessity for precise reading comprehension, logical interpretation, and the application of financial principles becomes evident. For instance, evaluating a company's operational status entails interpreting comprehensive financial statements to extract meaningful insights, identifying data patterns and relationships, and subsequently analyzing and formulating strategies. Furthermore, when making investment decisions, it is imperative to consider the cross-temporal and cross-domain characteristics of financial investments, conduct both fundamental and technical analyses, and select the optimal investment strategy amidst various uncertainties. Moreover, financial analysts need to navigate the complexities of financial regulations and compliance requirements.

In the realm of financial analysis, reasoning is of paramount importance. It involves utilizing available financial data, information, and pertinent factors to make judgments, draw conclusions, and infer insights about companies, businesses, projects, investments, or financial markets. This process demands critical thinking, analytical, and problem-solving skills. Financial reasoning further augments context and depth by considering broader economic, industry, and company-specific factors. Collectively, financial analysis and reasoning are indispensable for effective financial management and strategic planning, facilitating the examination, interpretation, and application of financial data to make well-informed decisions.

### 3.2. Rationale of Human Analysts and ChatGPT in Financial Analysis and Reasoning

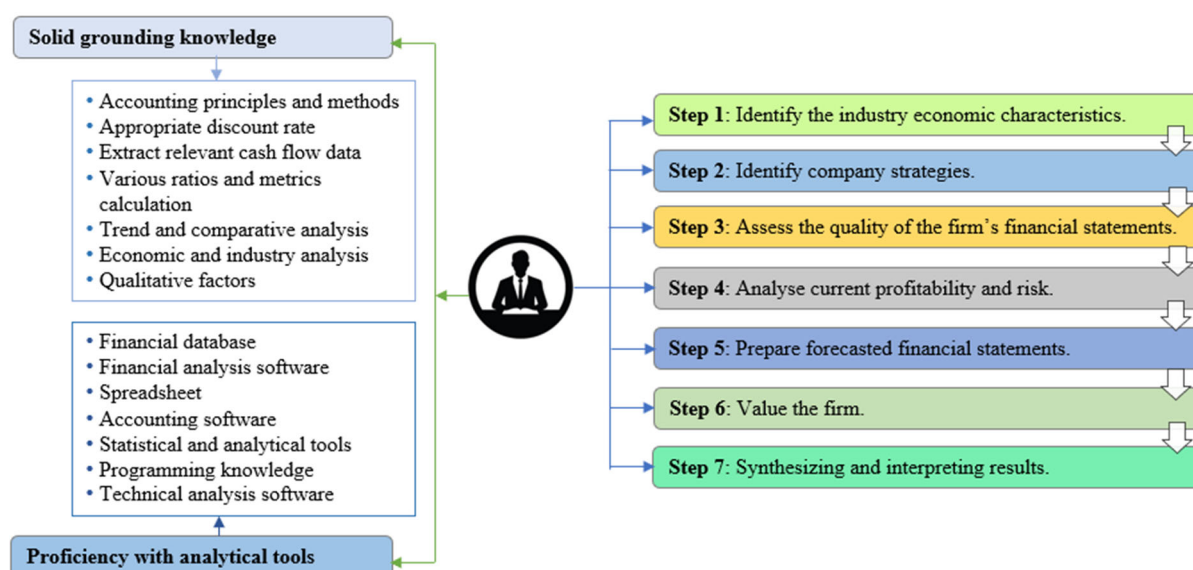
Financial professionals, including analysts, traders, and investors, typically engage in reasoning to scrutinize financial statements, evaluate performance metrics, forecast future outcomes, and formulate strategies for investment, budgeting, and financial planning. These professionals must have a solid foundation in algebra and mathematics to excel in their roles. They employ a range of sophisticated tools to support their research, analysis, and investment management endeavors, such as charting software, technical analysis applications, options and derivatives analyzers, portfolio management solutions,

and algorithmic trading platforms. Excel is fundamental for tasks like ratio analysis, risk management, investment analysis, and asset valuation. When managing extensive datasets, a deep understanding of mathematical and statistical techniques is crucial for drawing accurate conclusions from financial data. From simple to complex financial analysis, the following two example figures outline the process of analysis handled by human analysts. Figure 1 outlines a simple financial analysis process for net present value.



**Figure 1.** Steps for Calculating Net Present Value (NPV) by Human Analysts.

For complex financial analysis, such as effective financial statement analysis, financial analysts should possess a blend of knowledge and tools as outlined in Figure 2 (Masson 2018; Brealey et al. 2018; Wahlen et al. 2018):



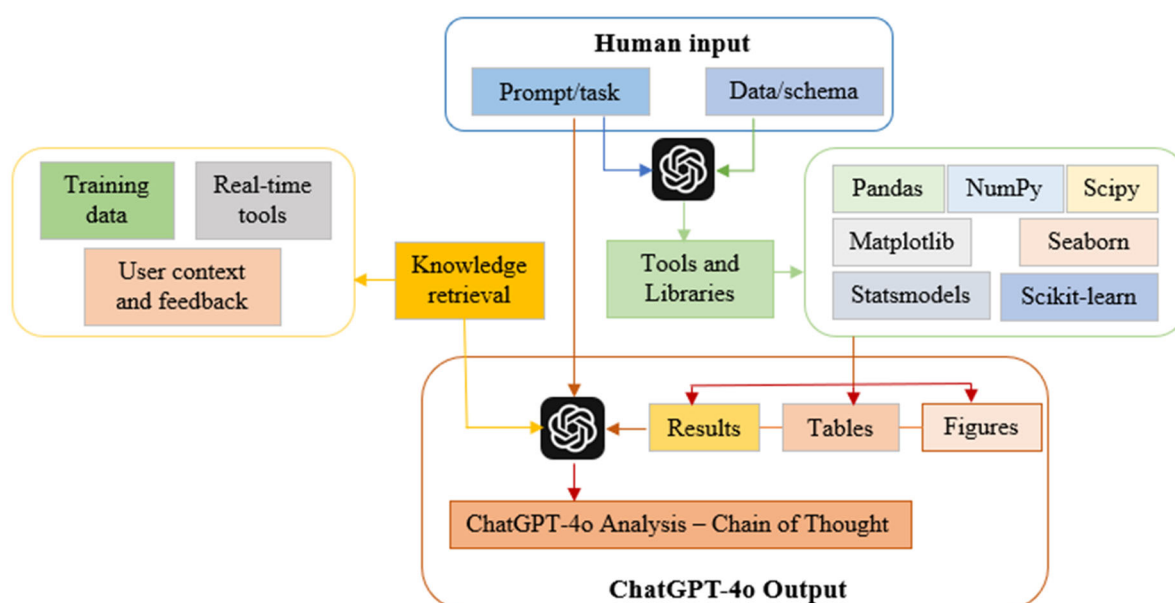
**Figure 2.** Steps for Financial Statement Analysis by Human Analysts.

Furthermore, these professionals are responsible for developing advanced financial models and conducting extensive research. Proficiency in specialized financial software and programming languages like C++, R, SAS, and Python is vital for effectively navigating the financial landscape.

Conversely, AI, a branch of computer science, focuses on developing systems and machines capable of performing tasks that typically require human intelligence, such as learning, reasoning, problem-solving, perception, language understanding, and decision-making (Sokolov 2019). Research by Son et al. (2023) on the application of large language models (LLMs) in financial reasoning confirms their capability to generate coherent investment opinions. Although this study does not detail the reasoning process in financial

analysis with LLMs, it underscores the importance of task formulation, synthetic data generation, prompting methods, and evaluation capability in influencing the quality of responses generated by LLMs. Complementing this, Wei et al. (2022) found that enabling a chain of thought or intermediate reasoning steps significantly enhances the complex reasoning capabilities of LLMs.

ChatGPT, a notable example of AI, exemplifies these capabilities, particularly in financial reasoning, demonstrating remarkable ability in complex multi-step reasoning tasks. Based on work by Cheng et al. (2023), Son et al. (2023), and Wei et al. (2022), we develop the following financial analysis and reasoning framework regarding ChatGPT-4o as a financial analyst, as outlined in Figure 3.



**Figure 3.** The Flow of Analysis for ChatGPT-4o as a Financial Analyst.

### 3.3. Tasks/Prompt

The principle guiding our task selection process is to ensure comprehensive coverage of financial concepts, the inclusion of realistic and complex task designs, and the integration of both quantitative and qualitative assessments. This approach equips AI models to effectively handle a wide range of financial analysis tasks in real-world settings.

We chose our tasks based on several key criteria to ensure the dataset's suitability for empirical testing in AI-driven financial analysis. Firstly, the diversity of financial scenarios, ranging from basic savings and investment calculations to complex option pricing and portfolio optimization, ensures exposure to various financial problems, enhancing AI model robustness and versatility. Secondly, the tasks are grounded in realistic financial activities, such as calculating future values, present values, and internal rates of return. These tasks mirror the analyses conducted by financial professionals, ensuring the dataset's relevance to practical applications.

Thirdly, we prioritized tasks that require complex reasoning and multi-step calculations, such as portfolio construction, capital budgeting analysis, and financial statement analysis. This complexity is ideal for testing AI systems' capabilities in handling sophisticated financial models and analyses. Additionally, the integration of various financial theories and models, including the Black–Scholes model for option pricing, the Gordon dividend model for valuing stocks, and the Modigliani–Miller theorem on capital structure, ensures a comprehensive understanding of financial principles.

Lastly, the tasks involve both quantitative calculations (e.g., yield to maturity) and qualitative assessments (e.g., financial leverage impact), which are crucial for developing AI that can interpret and analyze financial data effectively.

To gain deeper insights into the performance of ChatGPT-4o in managing these tasks, we have categorized them based on the complexity of the reasoning process into multi-step reasoning tasks and complex reasoning tasks. Additionally, we will assess the effectiveness of traditional tools used by human analysts, such as mathematical equations, Excel, Refinitiv, Stata, and other resources, as benchmarks to evaluate the achievement of our objectives.

### 3.3.1. Multi-Step Reasoning Tasks

The multi-step reasoning task consists of 32 questions covering various topics in corporate finance, investments, and derivatives. These topics include, but are not limited to, present value, future value, annuities, payment schedules, investment accumulation, and rate calculations. The task also explores basic futures and options pricing models, value calculation, risk management, and forecasting. Additionally, it includes qualitative assessments leading to decision-making or recommendations regarding future dividend payouts and capital structure. A detailed overview of these tasks can be found in Appendix A.

These tasks primarily require straightforward calculations or judgments involving a series of logical or computational steps to reach a specific conclusion. They are usually solvable through explicit logic and analysis without subjective judgments. It is expected that ChatGPT-4o will provide accurate computational formulas and resultant values when addressing such tasks. The aim is to evaluate ChatGPT's ability to apply logical and analytical reasoning in finance and investment, focusing on precision and objectivity in computations and assessments.

### 3.3.2. Complex Reasoning Tasks

Complex reasoning tasks require advanced calculations, extensive analysis, and creative thought processes, demanding a higher level of critical thinking compared to multi-step reasoning tasks.

To assess these analyses, we have developed six primary tasks. The first task evaluates ChatGPT's ability to perform technical analysis of randomly selected stocks and provide stock recommendations based on each technical indicator used. The second task aims to determine if ChatGPT can act as a portfolio manager by constructing an investment portfolio that meets the client's needs, with a focus on the application of Modern Portfolio Theory. The third task centers on corporate finance, emphasizing cash flow analysis and capital budgeting analysis. The fourth and fifth tasks are about financial statement analysis. The sixth task involves a binomial tree analysis. A detailed description of these tasks is available in Appendix B.

We conduct our evaluations using ChatGPT-4o, the latest and most advanced model equipped with a code interpreter and sophisticated data analysis capabilities. ChatGPT-4o excels in performing complex analyses and computations, allowing seamless interaction with various platforms and applications to ensure the accuracy and reliability of results. This enables comprehensive exploration and execution of tasks in finance and data analysis.

### 3.3.3. Evaluation Metrics

When financial analysts tackle a financial task, their approach typically begins with reasoning based on previously acquired specialized financial knowledge. They identify relevant concepts, formulas, and solutions applicable to the task at hand. Subsequently, they employ various professional tools to code and execute the task, culminating in the output of results. To scientifically compare the capabilities of large models like ChatGPT with traditional financial professionals, it is essential that these models also adopt a similar workflow. This workflow consists of logical reasoning followed by coding and modeling.

In evaluating the financial mathematics and decision-making performance of ChatGPT-4o, we will assess several metrics that encompass both quantitative and qualitative dimensions. These metrics are derived from generalized university rubrics, specifically tailored for elements of financial mathematics, designed to assess students' proficiency in comprehension, reasoning, modeling, data analysis, and critical thinking capabilities (Selke 2013). Consequently, we divide the key steps of task processing into two primary modules: reasoning and modeling. The reasoning module includes evaluative dimensions such as task understanding and task deconstruction, while the modeling module encompasses calculation ideas and formulas as well as accuracy. Additionally, we have incorporated an extra metric for critical thinking to assess ChatGPT-4o's ability in the application of knowledge and the level of critical thinking.

**Task Understanding:** This dimension gauges the ability to assimilate the prerequisites and objectives of a designated task or problem, evaluating the comprehension of the foundational concepts and principles inherent to the task.

**Task Deconstruction:** This dimension assesses the capability to fragment a task or problem into manageable and resolvable components or steps, focusing on the identification and isolation of pertinent variables and elements within a task.

**Calculation Ideas and Formulas:** This dimension scrutinizes the aptness and pertinence of the mathematical concepts, calculations, and formulas employed to decipher tasks, assessing the comprehension and application of mathematical models in problem resolution.

**Accuracy:** This metric quantifies the correctness and precision of the provided solutions against human analysts.

**Critical Thinking:** This dimension evaluates the capacity to objectively dissect information and formulate reasoned judgments, applying logical and reflective thinking to draw coherent conclusions and make informed decisions. The depth, quality, and efficacy of critical thinking can be assessed using diverse terminology that delineates the level of critical thinking applied (Stevens and Levi 2023).

For the criteria of task understanding, task deconstruction, and calculation ideas and formulas, we utilize qualitative scales categorized as basic, intermediate, and advanced to evaluate. The basic level identifies some components or steps of the task but lacks clarity and coherence in breaking it down and struggles to isolate pertinent variables and elements. The intermediate level represents effectively breaking down the task into clear, manageable components, accurately identifying and isolating pertinent variables and elements. The advanced level presents a skillful and coherent deconstruction of the task into detailed, manageable components, demonstrating precise identification and isolation of all pertinent variables and elements.

For the assessment of critical thinking/application of knowledge, we employ descriptors such as practical, applicable, functional, operational, and useful for questions 31 and 32 in the multi-step reasoning tasks. This practical descriptor evaluates if the knowledge applied is realistic and can be implemented in real-world scenarios. The applicable term assesses whether the knowledge is relevant and suitable for the given task. The functional descriptor evaluates if the applied knowledge effectively performs its intended purpose within the task. The operational descriptor checks if the knowledge can be actively used in real-world operations while considering all practical constraints and requirements. The useful descriptor measures the overall utility of the knowledge in achieving the task's objectives.

Conversely, for complex reasoning tasks such as investment suggestions and corporate strategy, the evaluative process is anchored in varying levels of critical thinking to appraise performance with terms including advanced, moderate, basic, superficial, and naive. The advanced level signifies a deep and thorough understanding, with the ability to analyze, synthesize, and evaluate information critically. It involves strategic thinking and insightful judgment. The moderate level indicates a reasonable level of critical thinking, where the individual can interpret and analyze information adequately but may not



demonstrate the same depth of insight as at the advanced level. The basic level shows a fundamental understanding and ability to apply critical thinking but with limited depth and complexity in reasoning. The superficial level suggests a shallow approach to critical thinking, where the individual's analysis and evaluation lack depth and are primarily surface-level. The naive level indicates a very simplistic and undeveloped approach to critical thinking, often characterized by a lack of understanding and basic reasoning skills.

This comprehensive evaluative framework ensures a nuanced and multifaceted assessment of both human analysts and ChatGPT in the domains of financial mathematics and decision-making. It allows for a robust comparison and analysis of competencies and proficiencies across diverse tasks and scenarios.

#### 4. Empirical Results and Findings

##### 4.1. Data Collection/Retrieval

First, it is evident that contemporary AI models, including those analogous to ChatGPT, lack the functionalities and capabilities for real-time data retrieval. Consequently, they cannot directly generate the datasets required for specific financial analyses. Instead, these models are primarily limited to guiding users on potential sources from which pertinent data can be acquired, as illustrated in Appendix C.

For academic pursuits, practitioners ranging from students to seasoned professionals such as analysts, traders, and investors might consider platforms like Yahoo Finance, which offers complimentary access to a vast array of financial data. However, for more comprehensive datasets, one may turn to institutional databases. Organizations often provide access to premium platforms like S&P Capital IQ, Bloomberg, and LSEG Refinitiv Workspace, among other specialized software, to facilitate in-depth financial analysis.

Consequently, the data used in our Complex reasoning tasks were sourced from S&P Capital IQ and LSEG Workspace for the following stocks listed on the ASX: Chalice Mining (CHN), Vulcan Energy Resources (VUL), Fineos Corporation (FCL), Southern Cross Gold Ltd. (SXG), Liontown Resources (LTR), Neuren Pharmaceuticals (NEU), WiseTech Global Ltd. (WTC), Aristocrat Leisure Limited (ALL), NextDC Ltd. (NXT), and Pro Medicus Limited (PME). Additionally, we retrieved Australian 10-year bond yields from Bloomberg on the 15th of May and divided it by 252 trading days to obtain the daily yield.

##### 4.2. Multi-step Reasoning Tasks Results and Findings

Based on the comprehensive multi-step analytical assessment presented in Table 1, it can be concluded that ChatGPT-4o demonstrates a proficient capability in basic or standard financial analysis reasoning. It follows a step-by-step procedure, working through sequential processes to find solutions akin to highly capable human analysts. In most cases (27 out of 30), ChatGPT-4o reaches accurate conclusions and exhibits a strong understanding of the task at hand. Due to space constraints, we are only displaying the task results that differ from those of human analysts. Results for other tests can be provided upon request.

**Table 1.** Multi-step reasoning task evaluation results for ChatGPT-4o. Task understanding, task deconstruction, and calculation ideas and formulas are evaluated with basic, intermediate, and advanced descriptors. Critical thinking/application of knowledge is evaluated with practical, applicable, functional, operational, and useful descriptors.

Task Number	Tasks	Task Understanding	Task Deconstruction	Calculation Ideas and Formulas	Accuracy	Critical Thinking/Application of Knowledge
1–8	Time value of money	advanced	advanced	advanced	Yes	functional
9	Investment yield	advanced	advanced	advanced	No	applicable
10–11	Effective rate, WACC	advanced	advanced	advanced	Yes	functional
12	Internal rate of return	advanced	advanced	advanced	No	applicable
13–18	Cost, valuation, option models	advanced	advanced	advanced	Yes	functional
19	Simple business valuation in M&A	advanced	advanced	advanced	No	applicable
20–30	Beta, bond, forecasting, pricing, arbitrage, risk, etc.	advanced	advanced	advanced	Yes	functional
31	Dividend payout suggestions	advanced	advanced	N/A	N/A	functional
32	Capital structure	advanced	advanced	N/A	N/A	functional

Several noteworthy insights emerge from the observations. Firstly, the importance of prompts cannot be overstated. Prompts are instructions or queries entered into the AI's interface to elicit responses, and they require careful wording and specific instruction. Inadequate instructions or poorly aligned Excel files often result in error messages and failure to achieve meaningful results. During our experiment, we observed that unclear instructions led to such issues.

Secondly, ChatGPT-4 demonstrates the ability to learn from instructions, supported by the study of Son et al. (2023), which shows that instruction-tuning plays a significant role in enhancing the performance of the model. Of the 30 calculation-focused multi-step reasoning tasks, the answers generated by ChatGPT-4o diverged from those provided by human analysts in only three instances: Tasks 9, 12, and 19. However, with the appropriate instructions or hints, ChatGPT-4o eventually arrives at the correct solutions, similar to those produced by skilled human analysts. For instance, in Task 9, ChatGPT-4o initially struggled with the exponential calculation, repeatedly arriving at an incorrect answer of 22.73%. After a question was asked, it corrected its answer to 19%, aligning with the human analysts' solution. However, on the following day, when the same question was asked again, ChatGPT-4o produced another incorrect answer by using a different approach. Detailed information can be found in Appendix D.

Task 12 involved calculating the internal rate of return (IRR). In the initial attempt, ChatGPT-4o employed a trial-and-error method but persisted in trying with larger rate numbers despite the net present value diminishing. A human analyst had to intervene and provide guidance, after which ChatGPT-4o completed the task. Subsequently, when the same task was entered again, ChatGPT-4o immediately produced the correct answer. However, on another fresh trial the next day, ChatGPT-4o generated an incorrect result by using Python. More detailed information is provided in Appendix E.

The issue with Task 19 pertained to the application of the weighted average cost of capital (WACC) for mergers and acquisitions (M&A). Initially, ChatGPT-4o incorrectly applied the WACC of the acquired firm, resulting in different outcomes compared to those of a human financial analyst. Upon receiving prompts about selecting the appropriate WACC for M&A, ChatGPT-4o correctly identified the use of the acquiring firm's WACC. Thus, with the proper instructions, it reached the correct conclusion.

Additional observations include instances where ChatGPT-4o does not directly provide final answers. In such cases, it recommends using tools like a financial calculator, Excel, or Python to complete the task.

For conceptual or qualitative tasks, such as Task 31 and Task 32, ChatGPT is capable of producing responses that are logical and adhere to recognized standards. However, these answers tend to be concise and may require further investigation. For example, in Task 31, which involves the understanding and insights into the dividend growth rate, ChatGPT-4o simply applied the average value, overlooking other elements that may affect the growth rate.

Moreover, it is noticeable that responses can vary each time a task is given, even though the main theme is maintained. This variability is a characteristic of artificial intelligence models. Language models, like chatbots, fundamentally operate as probabilistic systems, unlike deterministic systems. This means that posing the same questions can lead to different responses due to the inherent variability in the model's response generation. In these tasks, the wording and structure of the task significantly affect the resulting response generated by the model.

Conversely, for computational or quantitative tasks, the responses, including any incorrect outputs, tend to be consistent across multiple repetitions until intervention occurs. This consistency in computational tasks contrasts with the variability seen in responses to qualitative or conceptual tasks, underscoring the different response mechanisms inherent to artificial intelligence models in different task environments.

Overall, financial analysis is a critical task where even a small error can result in significant financial losses. The ongoing refinement and synergistic collaboration between LLMs and human expertise are crucial to melding analytical precision with human intuition. Therefore, it is recommended to utilize ChatGPT for analysis with great care and caution. It is imperative to always double-check the results to ensure accuracy.

#### 4.3. Complex Reasoning Tasks Results and Findings

In this section, we compare human analyst results with those produced by ChatGPT-4o for six complex reasoning tasks which cover the following broad areas: technical analysis and portfolio construction, capital budgeting and financial statement analysis, and derivatives option pricing. Table 2 presents the performance of ChatGPT-4o in executing complex reasoning tasks, highlighting its proficiency in answering these questions. For technical analysis and portfolio construction, we asked ChatGPT-4o to select the best 10 ASX-listed stocks based on the performance between January 2024 and the 14th of May 2024. Prompts and results from ChatGPT-4o are presented in Appendix F. After that, we extracted the daily stock prices from LSEG Workspace.

**Table 2.** Complex reasoning task evaluation results for ChatGPT-4o. Task understanding, task deconstruction, and calculation ideas and formulas are evaluated with basic, intermediate, and advanced descriptors. Critical thinking is evaluated with advanced, moderate, basic, superficial, and naïve descriptors.

Task Number	Tasks	Task Understanding	Task Deconstruction	Calculation Ideas and Formulas	Accuracy	Critical Thinking/Level of Critical Thinking
1.1	<b>Technical Analysis and Stock Recommendation</b> (Bollinger Bands)	advanced	advanced	advanced	Yes	advanced
1.2	Technical Analysis and Stock Recommendation (MACD)	advanced	advanced	advanced	Yes	advanced
1.3	Technical Analysis and Stock Recommendation (RSI)	advanced	advanced	advanced	Partially accurate	advanced
	<b>Portfolio Construction</b>					
2.1	Stock summary statistics	advanced	advanced	advanced	Yes	advanced
2.2	Correlation matrix	advanced	advanced	advanced	Yes	advanced

2.3	Portfolio Construction—Global Minimum Variance	advanced	advanced	advanced	Partially accurate	advanced
2.4	Portfolio Construction—Optimal Risky Portfolio	advanced	advanced	advanced	Yes	advanced
2.5	Efficient Frontier	advanced	advanced	advanced	Yes	advanced
3	Capital Budgeting	intermediate	basic	basic	No	naive
4	Financial Statement Analysis—Appendix B Q4	advanced	advanced	intermediate	Partially accurate	moderate
5	Financial Statement Analysis—Appendix B Q5	intermediate	intermediate	intermediate	No	superficial
6	Option pricing- Binomial Tree	advanced	advanced	moderate	No	moderate

For the technical analysis, particularly Tasks 1.1 to 1.3, ChatGPT-4o showed proficiency in performing tasks related to Bollinger Bands, Moving Average Convergence/Divergence (MACD), and Relative Strength Index (RSI). Following this, it offered individual stock recommendations—whether to buy, sell, or hold—based on the latest technical indicators available in our data sample. To validate the outcomes generated by ChatGPT-4o, we utilized LSEG/Refinitiv Workspace to create comparable results, typically formulated by us—human analysts. We included Chalice Mining Limited (CHN) as an example. As depicted in Table 3 and Appendix G, it was observed that the Bollinger Bands and MACD generated by ChatGPT-4o aligned with those from Workspace. However, discrepancies were identified in the RSI charts between ChatGPT-4o and Workspace.

**Table 3.** Complex Reasoning Task 1 demonstration: technical analyses and indicators for Chalice Mining Limited (CHN).

Human analyst result
Recommendations: Hold/Sell
Bollinger Bands: price is close to the upper band and a bullish reversal has recently occurred, suggesting a “hold” at this stage.
RSI: it is approaching the upper limit “70”, which may indicate “overbought” or “sell” situation.
MACD: MACD is above the signal line, indicating a “bullish” signal.
ChatGPT-4o result
“Recommendations: Hold/Sell
Overbought. MACD bullish. Price above upper Bollinger Band.”
Result comparison
Same recommendations, despite discrepancies in the RSI charts between ChatGPT-4o and LSEG Workspace.

Further, ChatGPT-4o demonstrated the capability to offer investment recommendations, providing rational justifications to back stock recommendations stemming from each technical indicator. For instance, it proposes a “hold/sell” recommendation when it detects a potential bullish crossover in the MACD when the RSI is approaching the upper limit and the price goes above the upper Bollinger band as the stock is in an overbought condition. Results for other stocks are available upon request.

For Complex Reasoning Task 2 (i.e., 2.1 to 2.5), ChatGPT-4o demonstrated proficiency in mirroring the responses of human analysts by constructing a global minimum variance portfolio and optimal risky portfolio, determining the weights of each stock in the portfolios, and combining the portfolios. However, there is a discrepancy in the stock weights of the global minimum variance portfolio determined by Excel/Stata and ChatGPT-4o, as shown in Table 4. For an optimal risky portfolio, stock weights provided by ChatGPT-4o are almost the same as the weights computed by Excel and Stata. ChatGPT-4o also successfully constructed an efficient frontier promptly. Both ChatGPT-4o’s calculations and our analyses are also presented in Appendix H.

**Table 4.** Complex Reasoning Task 2 demonstration: portfolio construction.

Human analyst result
Global minimum variance portfolio weights based on Excel and Stata: CHN: 0; VUL: 1.65%; FCL: 5.86%; SXG: 2.19%; LTR: 1%; NEU: 8.69%; WTC: 5.15%; ALL: 56.61%; NXT: 13.46%; PME: 5.49%
Optimal risky portfolio weights based on Excel and Stata: CHN: 0; VUL: 6.34%; FCL: 0; SXG: 18.40%; LTR: 0; NEU: 0; WTC: 25.39%; ALL: 0; NXT: 34.86%; PME: 15.01%
ChatGPT-4o result
Global minimum variance portfolio weights: “CHN: 0; VUL: 4.83%; FCL: 10.94%; SXG: 4.94%; LTR: 4.69%; NEU: 14.50%; WTC: 14.97%; ALL: 16.20%; NXT: 13.83%; PME: 15.11%”
Optimal risky portfolio weights: “CHN: 0; VUL: 6.37%; FCL: 0; SXG: 18.40%; LTR: 0; NEU: 0; WTC: 25.44%; ALL: 0; NXT: 34.85%; PME: 14.94%”
Result comparison
ChatGPT-4o and Excel/Stata generated similar weights for optimal risky portfolios; however, the weights generated by ChatGPT-4o for the global minimum variance portfolio are different from the results produced by Excel and Stata.

However, ChatGPT-4o faced challenges in completing Complex Reasoning Tasks 3, 4, and 5. Task 3 assessed ChatGPT-4o’s ability in capital budgeting analysis, evaluating its proficiency in interpreting extensive information, distinguishing relevant information, and critical thinking. The results provided in Table 5 show that ChatGPT-4o’s final answers for NPV were inconsistent with human analyst calculations. Appendix I further shows errors in analyzing information, recognizing irrelevant costs, and miscomputing depreciated expenses. Moreover, ChatGPT-4o failed to create a detailed capital budgeting template outlining each cash inflow and outflow item annually.

**Table 5.** Complex Reasoning Task 3 demonstration: capital budgeting.

Human analyst result
At 12% discount rate: NPV = \$255,234.67
At 14% discount rate: NPV = \$115,539.10
ChatGPT-4o result
At 12% discount rate: “The NPV of the project is \$1,071,417.83, indicating that the project is profitable.”
At 14% discount rate: “The NPV of the project is \$966,137.31, still indicating profitability, but with a reduced margin due to the higher risk reflected in the higher discount rate.”
Result comparison
NPVs computed by ChatGPT-4o are completely different from the correct answers.

In Tasks 4 and 5, we asked ChatGPT-4o to conduct financial statement analyses (Task 4 is a basic financial statement analysis, and Task 5 is a complex financial statement analysis). These tasks were sourced from the CFA problems test bank in the book *Essentials of Investments* (Bodie et al. 2022). However, the results varied significantly when compared to those of a human analyst. For example, without explicit instruction, ChatGPT-4o would apply the three-component DuPont formula analysis for Task 4 instead of the commonly used five-component method (Table 6). For task 5, ChatGPT-4o computed DuPont components incorrectly (Table 7). Step-by-step calculations from ChatGPT-4o are presented in Appendices J and K. Additionally, it appears that ChatGPT-4o struggles to accurately retrieve data tables formatted as images. The Excel template created by ChatGPT-4o displays different values. Any issues encountered in the initial step led to markedly different results or interpretations in subsequent steps.

**Table 6.** Complex Reasoning Task 4 demonstration: financial statement analysis.

Human analyst result
Part a: components of the DuPont formula
Tax burden: 0.6335
Interest burden: 0.9699
Profit margin: 0.1615
Asset turnover: 1.6992
Leverage: 1.4070
Part b: ROE:
ROE = 23.72%
Part c: growth rate
$g = 16.46\%$
ChatGPT-4o result
Part a: components of the DuPont formula
“Net profit margin: 0.099
Asset turnover: 1.658
Equity multiplier: 1.409”
Part b: ROE:
“ROE = 23.2%”
Part c: growth rate
“ $g = 16.3\%$ ”
Result comparison
The ROE and growth rates calculated by ChatGPT-4o are correct. However, ChatGPT-4o only performed 3-step DuPont calculations. It requires further instruction to conduct 5-step DuPont analyses.

**Table 7.** Complex Reasoning Task 5 demonstration: financial statement analysis.

Human analyst result
Part a–i: components of the DuPont formula in 2022
White Brush (ProtectBrush)
Tax burden: 0.6744 (0.6599)
Interest burden: 1.00 (0.9545)
Profit margin: 0.0851 (0.0646)
Asset turnover: 1.8259 (3.6286)
Leverage: 1.5071 (1.5386)
ROE: 15.8% (22.7%)
Parts a–ii: ROE and plowback in 2022
White Brush (ProtectBrush)
ROE = 15.8% (22.7%)
Plowback = 1.00 (0.344)
Sustainable growth rate = 15.8% (7.8%)
Part b: explanation of WhiteBrush’s EPS growth rate
WhiteBrush’s recent EPS growth has been achieved by increasing book value per share, not by achieving greater profits per dollar of equity. Since EPS is equal to (Book value per share $\times$ ROE), a firm can increase EPS even if the ROE is declining; this is the case with WhiteBrush. WhiteBrush’s book value per share has more than doubled in the last two years. Book value per share can increase either by retaining earnings or by issuing new stock at a market price greater than book value. WhiteBrush has been retaining all earnings, but the increase in the number of outstanding shares indicates that it has also issued a substantial amount of stock.
ChatGPT-4o result
Parts a–i: components of the DuPont formula in 2022
WhiteBrush (ProtectBrush)
“Net profit margin: 0.0573 (0.0373)
Asset turnover: 1.42 (3.61)
Equity multiplier: 3.12 (3.30)
ROE: 25.38% (44.51%)”
Parts a–ii: ROE and plowback in 2022
White Brush (ProtectBrush)
“ROE = 25.38% (44.51%)

Plowback = 0.75 (-0.25)*	
Sustainable growth rate = 19.04% (-11.13%)”	
*“Dividend payout ratio is 125%. ProtectBrush is paying out more than its earnings in dividends, resulting in a negative plowback ratio.”	
Part b: explanation of WhiteBrush’ EPS growth rate	
“WhiteBrush has produced an average annual EPS growth rate of 40% over the last two years despite a declining ROE. Possible reasons include:	
1.	Revenue Growth: WhiteBrush has shown substantial revenue growth (from \$3480 in 2020 to \$7760 in 2022), which has increased the overall earnings.
2.	Cost Management: The company has managed its costs effectively, maintaining a relatively stable profit margin.
3.	Share Buybacks: The average shares outstanding increased marginally, suggesting that the company might be using some of its earnings to repurchase shares, thereby boosting EPS.
4.	Dividend Payout: A lower dividend payout ratio in recent years indicates that more earnings are retained within the company, contributing to EPS growth.
These factors combined can lead to a higher EPS even if the ROE is declining.”	
Result comparison	
In parts a–i, although instruction has been given for 5-step DuPont analyses, ChatGPT-4o was only able to demonstrate 3-step DuPont calculations. The components were computed incorrectly in both parts a–i and a–ii. In part b, ChatGPT-4o was able to provide superficial discussions. It discussed share buyback. However, WhiteBrush’s shares outstanding increased from 2021 to 2022.	

Task 6 assessed whether ChatGPT-4o could calculate the American call option price using the binomial tree approach. As shown in Table 8, ChatGPT-4o concluded that early exercise is not optimal at any step, which is not correct in view of the output from DerivaGem. Appendix L further shows that ChatGPT-4o could not display the binomial tree diagram, even after instructing it to follow the DerivaGem diagram. Lastly, we attempted to use the new voice interaction feature in ChatGPT-4o. It could provide a better tree diagram, but the option prices and early exercise decisions remained incorrect.

**Table 8.** Complex Reasoning Task 6 demonstration: Option Pricing–Binomial Tree.

Human analyst result	
DerivaGem output shows that early exercise is optimal at node D with a value of 0.0538. The value of the option at the initial node A is 0.0188.	
ChatGPT-4o result	
“There are no nodes where early exercise would be optimal since the option values are non-negative and less than the intrinsic values at every point. Therefore, early exercise is not optimal at any step.”	
The value of the option at the initial node A computed by ChatGPT-4o is 0.0777.	
The value of the option at the initial node A computed based on ChatGPT-4o voice interaction is 0.0255.	
Result comparison	
The values of the option computed by ChatGPT-4o were incorrect. In addition, its decision on early exercise was incorrect.	

During our complex reasoning evaluation, several issues related to ChatGPT-4o were identified. First, even when the same methods have been applied, a discrepancy exists between the charts produced by ChatGPT-4o and Workspace. Since both ChatGPT-4o and Workspace are tools or software used by human analysts to draw conclusions, it is plausible that the charts are slightly different from one another. Despite the existing discrepancies in both charts, the stock recommendations using the RSI from both ChatGPT-4o and Workspace are consistent (the RSI value lies between lower and upper bands).

Second, ChatGPT-4o relies mainly on Python programming. According to Dilmegani (2024), “the code interpreter only supports Python as a language”. Differences in programming methods may cause differences, such as the stock weights in the construction of a global minimum variance portfolio. In addition, this requires Python experts or analysts with proficient Python skills to detect any discrepancies in the calculation method.

Third, the capital budgeting and financial statement analyses exposed a shortfall in ChatGPT-4o’s capability to replicate human analytical processes, particularly in offering

sequential calculations and in creating Excel-like templates outlining each cash flow item. This indicates that ChatGPT-4o generates responses based on patterns learned during training and does not understand context or infer meanings in the way humans do. This highlights a limitation in ChatGPT-4o's ability to accurately process comprehensive information, suggesting a potential obstacle in its capability to assimilate and analyze complex data sets accurately.

Lastly, ChatGPT-4o may not provide accurate results for specialized finance areas such as derivative securities. Although GPT-4o was able to perform the step-by-step calculations like a human analyst, the results, such as those involving the probability of the up move, were not correct. Furthermore, it has to rely on Python programming to display the tree diagram, but the structure is somewhat different from a normal binomial tree diagram. The new voice and video model introduced by OpenAI on 14 May 2024 was able to generate a better tree diagram; however, the value of the option computed was also incorrect.

#### *4.4. Discussion and Practical Application and Implementation*

Consistent with the findings in a study from Cheng et al. (2023), ChatGPT-4o is able to achieve comparable performance to human analysts, at least for entry-level analysts. Our findings also align with Kocóń et al. (2023), which showed that the more difficult the task, the higher the performance loss of ChatGPT-4o. The qualitative analysis revealed ChatGPT's lack of deep thinking and comprehensive analysis. Our results provide a basis for a fundamental discussion on whether high-quality financial analysis and reasoning can be effectively applied in real-life scenarios.

First of all, to understand whether ChatGPT provided the correct answers, users should have possessed enough prerequisite knowledge. Second, ChatGPT-4o enhances financial analysis efficiency by performing both basic and complex tasks, thereby automating repetitive calculations and allowing financial analysts to concentrate on more strategic decision-making. Financial institutions can deploy ChatGPT-4o to handle routine tasks such as present value calculations, ratio analysis, and basic forecasting, thus streamlining operations and optimizing human resource utilization. Additionally, ChatGPT-4o contributes to cost reduction by automating numerous financial analysis processes, which is particularly advantageous for small and medium-sized enterprises that may lack extensive financial analysis teams. In the realm of investment strategies, ChatGPT-4o's ability to conduct technical analysis and portfolio construction enables it to assist in developing and optimizing investment strategies including analyzing stock performance, recommending buy/sell/hold actions, and constructing diversified portfolios based on Modern Portfolio Theory. Furthermore, ChatGPT-4o serves as a valuable educational tool, offering finance students and professionals step-by-step explanations and analyses of various financial concepts and tasks, thus aiding in the comprehension of complex financial models and theories.

However, given the limitations of ChatGPT now, several issues should be kept in mind. Firstly, data accuracy and reliability can be a concern, as ChatGPT might sometimes provide incorrect or outdated information based on its training data. Secondly, contextual understanding can be limited, with the AI potentially misinterpreting complex financial scenarios or nuances that a human analyst would catch. Thirdly, dependence on input quality is crucial; the outputs generated by ChatGPT are only as good as the data and queries it receives, necessitating careful and precise input from users. Fourthly, a lack of real-time updates means that ChatGPT cannot access the latest data or trends beyond its training cutoff, limiting its usefulness for dynamic, real-time financial analysis. Fifthly, security and privacy are important considerations, as using AI for financial analysis involves handling sensitive financial data, requiring robust measures to protect against data breaches. Lastly, ethical considerations arise from the potential biases inherent in AI models, which can impact the fairness and objectivity of the analysis. Addressing these issues



is essential for effectively leveraging ChatGPT in financial analysis while mitigating potential risks.

## 5. Conclusions

This study has examined the analytical and reasoning capabilities of ChatGPT-4o through various financial tasks, providing significant insights into the strengths and limitations of LLMs in financial analysis. ChatGPT-4o has demonstrated considerable skill in performing standard financial reasoning tasks, closely aligning its analytical approach with that of human analysts. It excels in logical reasoning, task decomposition, and generating solutions, which are essential for tasks like financial modeling and forecasting. However, the study also highlights several challenges and limitations.

The variability in ChatGPT-4o's responses, especially for qualitative tasks, underscores the importance of explicit instructions and careful task formulation. The discrepancies observed in some tasks between ChatGPT-4o and human analysts emphasize the need for robust evaluation metrics to ensure consistent and reliable outputs. Additionally, ChatGPT-4o encountered difficulties with complex tasks requiring a higher level of analytical depth and comprehensive understanding, indicating its limitations in replicating intricate human analytical methods.

Despite these challenges, the prospective integration of ChatGPT-4o with specialized financial data providers and tools, such as Bloomberg, S&P Capital IQ, and statistic packages, represents a transformative shift in the financial sector. This integration is poised to significantly enhance human analytical processes, enabling financial professionals to concentrate more on critical decision-making elements.

This research contributes to the understanding of AI's role in finance by providing detailed insights into the applications and limitations of ChatGPT-4o in financial analysis. It establishes that while ChatGPT-4o can effectively perform basic and some complex financial tasks, it struggles with tasks requiring deep analytical depth and critical thinking. The study's findings enhance the knowledge base for academicians, developers, and stakeholders interested in integrating AI into financial practices, demonstrating the potential for AI to enhance efficiency and accuracy in financial analysis when combined with human expertise.

This study is not without its limitations. While the tasks tested are grounded in real-life scenarios, there is a need to incorporate more high-level, practical, and specific tasks to further evaluate the capabilities of AI models. Expanding the dataset to include these advanced tasks will provide a more rigorous assessment of the models' performance in complex financial environments. Furthermore, the study utilized only one AI model, ChatGPT-4o. Future research should consider using a variety of AI models, such as LLaMA, Galactica, and Pythis, including those developed by specific financial firms, to enable comprehensive comparisons and determine which models produce the most accurate and reliable results.

Additionally, the study assumed a general classification of human analysts as senior and expert analysts. However, human analysts vary widely in expertise, including junior, mid-level, and high-level analysts. Identifying and incorporating specific levels of human analysts in future evaluations could provide deeper insights and more nuanced comparisons of AI model performance against varying levels of human expertise.

Future research should focus on enhancing the deep thinking and comprehensive analysis capabilities of AI models like ChatGPT-4o. This could involve the development of hybrid models that combine the strengths of AI and human intelligence, leveraging AI's computational power and efficiency with human intuition and contextual understanding. Real-time data integration and continuous learning mechanisms could be explored to improve AI's adaptability to dynamic financial environments. Additionally, ethical considerations and overcoming biases in AI models should be a priority, ensuring fair and objective financial analysis. Continued interdisciplinary research will be essential to fully realize the potential of AI in finance.

**Author Contributions:** Conceptualization, L.X.L.; methodology, L.X.L. and K.X.; validation, L.X.L. and C.C.; formal analysis, L.X.L. and Z.S.; data curation, L.X.L. and Z.S.; writing—original draft preparation, writing—review and editing, and supervision, L.X.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The raw data and the detailed empirical results are available from the corresponding author on request.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Multi-Step Reasoning Tasks/Prompt

1. Suppose you deposit \$1000 in a savings account that pays 10% interest, compounded quarterly. How much will be in that account after 10 years if there is no withdrawal?
2. Sammy deposits \$1000 now, \$1500 in one more year, then \$2000 in two years, and \$2500 in three years in a savings account that pays 10% interest per annum. How much does Sammy have in the account at the end of the third year?
3. You will deposit \$1500 in one year's time from now, \$2000 in two years' time, and \$2500 in three years' time, in an account paying 10 percent interest per annum. What is the present value of these cash flows?
4. You are purchasing a home and are scheduled to make 30 annual installments of \$10,000 per year. Given an interest rate of 5%, what is the price you are paying for your house?
5. The superannuation guarantee rate in the industry is 9.5% in Australia. If your annual income is \$100,000, you will have \$9500 every year for the next 30 years till your retirement (ignore the growth of income here). Given a 10% rate of interest, how much will you have saved by the time you retire?
6. Suppose you are valuing an investment that promises \$100 per year at the end of this and the next four years. If the annual interest rate is 10%, calculate the value of this investment.
7. You wish to invest in financial security with a face value of \$500,000, a term to maturity of 180 days, and a yield of 8.75% per annum. How much will it cost you today to buy?
8. As a winner of a dragon boat competition, you can choose one of the following prizes:
  - a. \$100,000 now.
  - b. \$180,000 at the end of five years.
  - c. \$11,400 a year forever.
  - d. D. \$19,000 for each of 10 years.
  - e. \$6,500 next year and increasing thereafter by 5% a year forever.

Assume the interest rate is 12%. What is your choice?

9. Google Inc. became a public company when it conducted an IPO of ordinary shares in August 2004. It was originally priced at \$85 per share. By August 2018, Google shares stood at \$1084. What annual rate of return did the investors who bought Google shares at the IPO and held them until August 2018 earn?
10. It is 30 June. ABC company has a commercial bill that has a current interest rate yield of 6.08 percent per annum. The existing bills mature on 31 August 2015 but will be replaced by a further issue at that date. What is the effective annual interest rate on the bill?
11. What is the WACC for a firm with \$30 million in outstanding debt with a required return of 8%, 8 million in equity shares outstanding trading at \$15 each with a required return of 12%, and a tax rate of 35%?
12. Consider an investment that costs \$800 and has cash flows of 300, 200, 150, 122, and 133 in years 1–5. Calculate the internal rate of return.
13. ABC Corporation has a stock price of \$50. The firm has just paid a dividend of \$3 per share, and shareholders think that this dividend will grow by a rate of 5% per year. Use the Gordon dividend model to calculate the cost of equity for ABC.
14. ABC Corporation has just paid a dividend of \$3 per share. You, an experienced analyst, feel quite sure that the growth rate of the company's dividends over the next ten years will be 15% per year. After ten years, you think that the company's dividend growth rate will slow to the industry average, which is about 5% per year. If the cost of equity for ABC is 12%, what is the value today of one share of the company?
15. Your firm, ABC, is considering acquiring a business. Calculate its value using the following information: Firm ABC's WACC is 12.5%, and the cash flows of the business are \$1 million for years 1–4. The business is expected to grow at a rate of 5% after the fourth year.
16. The current level of the S&P 500 is 3000. The dividend yield on the S&P 500 is 2%. The risk-free interest rate is 1%. What should be the price of a one-year maturity futures contract?
17. A stock selling for \$25 today will, in one year, be worth either \$35 or \$20. If the interest rate is 8%, what is the value today of a one-year call option on the stock with an exercise price of \$30? Use the simultaneous equation approach to price the option.

18. Use the Black–Scholes model to price a call option on a stock whose current price is 50, with an exercise price of 50, an interest rate of 10%, a maturity of 0.5 years, and a standard deviation of 25%.
  19. You are analyzing Woolworth's potential acquisition of Billabong. Suppose Woolworths plans to offer \$450 million as the purchase price for Billabong, and it will need to issue additional debt and equity to finance the acquisition. You estimate that the issuance costs will be \$15 million and will be paid as soon as the transaction closes. You estimate the incremental free cash flows from the acquisition will be \$29 million in the first year and will grow at 4% per year thereafter. What is the NPV of the proposed acquisition? You may access the other information from the file uploaded.
  20. The spreadsheet uploaded is the five-year monthly prices for Intel Corporation and the S&P 500. Calculate Intel's beta.
  21. ABC Corporation has issued 1 million fully paid ordinary shares. The after-tax profits for ABC are \$500,000. Earnings per share are 50 cents ( $\$500,000/1\text{ m} = 50\text{ cents}$ ). ABC's shares are currently selling at a price–earnings multiple of 10. The financial manager of ABC is planning a 1-for-5 bonus issue. Answer the following questions.
    - a. What is the current share price?
    - b. How many new shares will be issued under the 1-for-5 bonus scheme?
    - c. What are the new earnings per share after the bonus issue?
    - d. What is the market price after the bonus issue if the price–earnings multiple remains at 10?
    - e. After the bonus issue, what is the total value of the investor's holdings? Assume this investor previously had 10 shares.
  22. You looked at the newspaper quotes for options on ABC share; you saw that a March call option with a strike price of 37.5 is priced at 6.375, whereas the May call option with the same exercise price is priced at 6.
    - a. Can you devise an arbitrage out of these prices?
    - b. Do you have an explanation for the newspaper quotes?
  23. An American call option is written on a stock whose price today is \$60. The exercise price of the call is \$45.
    - a. If the call price is 2, explain how you would use arbitrage to make an immediate profit.
    - b. If the option is exercisable at time  $T = 1$  year, and if the interest rate is 10%, what is the minimum price of the option? Use proposition a.
  24. A one-year gold futures contract is selling for \$1558. Spot gold prices are \$1500, and the one-year risk-free rate is 4%.
    - a. According to spot-futures parity, what should the futures price be?
    - b. What risk-free strategy can investors use to take advantage of the futures mispricing, and what will be the profits of the strategy?
  25. Based on the monthly TV Ads released and the revenues recorded over the past one-year period, forecast next year's revenue. And if there are 100 Ads put on next month, what is next month's revenue? You may access the data from the file uploaded.
  26. Suppose a fund manager has a portfolio that consists of a single asset. The return of the asset is normally distributed with a mean return of 20% and a standard deviation of 30%. The value of the portfolio today is \$100 million.
    - a. What is the distribution of the end-of-year portfolio value?
    - b. What is the probability of a loss of more than \$20 million by year-end? For example, what is the probability that the end-of-year value is less than \$80 million?
    - c. With a 1% probability, what is the maximum loss at the end of the year?
  27. Your company is considering either purchasing or leasing an asset that costs \$1,000,000. The asset, if purchased, will be depreciated on a straight-line basis over six years to a zero residual value. A leasing company is willing to lease the asset for \$300,000 per year; the first payment on the lease is due at the time the lease is undertaken (i.e., year 0), and the remaining five payments are due at the beginning of years 1–5. Your company has a tax rate of 40% and can borrow at 10% from its bank.
    - a. Should your company lease or purchase the asset?
    - b. What is the maximum lease payment it will agree to pay?
  28. A one-year, \$100,000 loan carries a coupon rate and a market interest rate of 12%. The loan requires payment of accrued interest and one-half of the principal at the end of six months. The remaining principal and accrued interest are due at the end of the year. What is the duration of this loan?
  29. On 23 January 1999, the market price of a Bond was \$1122.32. The bond pays \$59 in interest on 1 March and 1 September of each of the years 1999–2005. On 1 September 2005, the bond was redeemed at its face value of \$1000.
    - a. Calculate the yield to maturity of the bond.
    - b. Calculate its duration.
  30. An investment fund owns the following portfolio of three fixed-rate government bonds: You may access the data from the file uploaded.
- The total market value of the portfolio is US\$96,437,017. Each bond is on a coupon date so that there is no accrued interest. The market values are the full prices given the par value. Coupons are paid half yearly. The yields to maturity are stated for a periodicity of 2. The Macaulay durations are annualized.
- a. Calculate the average (annual) modified duration for the portfolio using the shares of market value as the weights.
  - b. Estimate the percentage loss in the portfolio's market value if the annual yield to maturity on each bond goes up by 20 bps.

31. BHP in Australia is still generating good profits. But growth is slowing down. Based on BHP's previous 10-year dividend payout history, help the CFO decide how to start up a program for paying out cash to stockholders. Access the data from the file uploaded.
32. A few years after being appointed financial manager at Sedona Fabricators, Inc., you are asked by your boss to prepare for your first presentation to the Board of Directors. This presentation will pertain to issues associated with capital structure. It is intended to ensure that some of the newly appointed, independent board members understand certain terminology and issues. As a guideline for your presentation, you are provided with the following outline of questions.
  - a. What is capital structure?
  - b. What is financial leverage?
  - c. How does financial leverage relate to company risk and expected returns?
  - d. Modigliani and Miller demonstrated that capital structure policy is irrelevant. What is the basis for their argument? What are their Propositions I and II?
  - e. How does the introduction of corporate taxes affect the M&M model?
  - f. How do the costs of insolvency and financial distress affect the M&M model?
  - g. What are agency costs? How can the use of debt reduce agency costs associated with equity?

### Appendix B. Complex Reasoning Tasks/Prompt

1. Using the historical price data for the 10 stocks named CHN, VUL, FCL, SXG, LTR, NEU, WTC, ALL, NXT, and PME in the attached file, please conduct technical analyses using the following indicators: Bollinger Bands, Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD).
  - a. Please draw technical charts and indicators for each stock.
  - b. Assume today is the 15th of May 2024. What is your overall recommendation for each stock based on the charts and indicators?
  - c. Please summarise the recommendations in a table with the explanation provided.
2. Using the historical price data for the 10 stocks named CHN, VUL, FCL, SXG, LTR, NEU, WTC, ALL, NXT, and PME in the attached file, please first sort the date from the earliest to the latest date and then perform the following analyses. Please save the return data in the same excel file:
  - a. Please calculate daily return and then present summary statistics for each stock in a table such as Mean Return, Standard Deviation, Max, Min, Median, Skewness, and Kurtosis in a table. Then, discuss whether the returns of each stock follow normal distribution based on summary statistics.
  - b. Please display the correlation matrix based on the returns for the 10 stocks and indicate whether it is significant at 1%, 5%, and 10% levels using \*\*\*, \*\*, \* respectively. Please discuss the results and significance of the correlations between different pairs of stocks.
  - c. Based on the average returns, standard deviations, and correlation matrix for the 10 stocks, please construct a global minimum variance portfolio. What criterion did you consider when creating the global minimum variance portfolio? Assume the risk-free rate is 0.017% and there is no short selling in any stock, all weights should lie between 0 and 1. Please calculate and explain the weights of stocks in the global minimum variance portfolio. Based on the weights you have calculated, what is the global minimum variance portfolio return and standard deviation?
  - d. Based on the average returns, standard deviations, and correlation matrix for the 10 stocks, please construct an optimal risky portfolio. What criterion have you considered when creating the optimal risky portfolio? Assume the risk-free rate is 0.017% and there is no short selling in any stock, all weights should lie between 0 and 1. Please calculate and explain the optimal weights of stocks in the optimal risky portfolio. Based on the weights you have calculated, what is the optimal risky portfolio return and standard deviation?
  - e. Please create an efficient frontier for the combinations of these 10 stocks and also indicate the global minimum variance and optimal risky portfolios on the graph.
3. Aus Car Execs (ACE) is set up as a sole trader and is analyzing whether to enter the discount used rental car market. This project would involve the purchase of 100 used, late-model, mid-sized automobiles at the price of \$9500 each. In order to reduce their insurance costs, ACE will have a LoJack Stolen Vehicle Recovery System installed in each automobile at a cost of \$1000 per vehicle. ACE will also utilize one of its abandoned lots to store the vehicles. If ACE does not undertake this project, they could sublease this lot to an auto repair company for \$80,000 per year. The \$20,000 annual maintenance cost on this lot will be paid by ACE, whether the lot is subleased or used for this project. In addition, if this project is undertaken, net working capital will increase by \$50,000.
 

For taxation purposes, the useful life of the automobiles is determined to be 5 years, and they will be depreciated using the diminishing value method. Each car is expected to generate \$4800 a year in revenue and have operating costs of \$1000 per year. Starting 6 years from now, one-quarter of the fleet is expected to be replaced every year with a similar fleet of used cars. This is expected to result in a net cash flow (including acquisition costs) of \$100,000 per year continuing indefinitely. This discount rental car business is expected to have a minimum impact on ACE's regular rental car business, where the net cash flow is expected to fall by only \$25,000 per year. ACE expects to have a marginal tax rate of 32%.

Based on the above information, if ACE uses a discount rate of 12% for capital budgeting, what is the NPV of this project? If ACE adjusts the discount rate to 14% to reflect higher project risk, what is the NPV? For each question, please construct a capital budgeting analysis.

4. John is reviewing ABC's financial statements to estimate its sustainable growth rate. Using the information presented in the Table uploaded, can you please first convert this into an Excel template (save it as a separate file) and (measurement: \$ million, except per-share data)?
  - i. Identify and calculate the components of the DuPont formula.
  - ii. Calculate the ROE (Return on Equity) for 2022 using the components of the DuPont formula.
  - iii. Calculate the sustainable growth rate for 2022 from the firm's ROE and plowback ratio. (Bodie et al. 2022, p. 468)
5. Jennifer is a recently hired analyst. After describing the electric toothbrush industry, her first report focuses on two companies, WhiteBrush company, and ProtectBrush company, and concludes:
 

WhiteBrush is a more profitable company than ProtectBrush, as indicated by the 40% sales growth and substantially higher margins it has produced over the last few years. ProtectBrush's sales and earnings are growing at a 10% rate and produce much lower margins. We do not think ProtectBrush is capable of growing faster than its recent growth rate of 10%, whereas WhiteBrush can sustain a 30% long-term growth rate. Please convert the information in the screenshots into an Excel template and save it as a separate file.

  - (a) Criticize Jennifer's analysis and conclusion that WhiteBrush is more profitable, as defined by return on equity (ROE), than ProtectBrush and that it has a higher sustainable growth rate. Use only the information provided in Table WhiteBrush and Table ProtectBrush. Support your criticism by calculating and analyzing:
    - i. The five components that determine ROE.
    - ii. The two ratios that determine sustainable growth: ROE and plowback.
  - (b) Explain how WhiteBrush has produced an average annual earnings per share (EPS) growth rate of 40% over the last two years with an ROE that has been declining. Use only the information provided in Table WhiteBrush. (Bodie et al. 2022, p. 468)
6. The Australian dollar is currently worth USD 0.6100, and this exchange rate has a volatility of 12%. The Australian risk-free rate is 7%, and the U.S. risk-free rate is 5%. Use a three-step binomial tree to value a 3-month American call option with a strike price of USD 0.6000. Please draw a tree diagram and show the value of currency and value of option at each node in the diagram and also indicate whether early exercise is optimal (in red colour). Assume the U.S. is a domestic country and Australia is a foreign country (Hull 2015, Example 13.2, p. 314).

## Appendix C. Data Collection/Retrieval



To provide the specific historical monthly price data for the requested stocks, I will need to use a data source that offers this information. While I cannot generate this data directly from within my current environment, I can guide you on how to obtain this data from reliable financial data sources.

### Steps to Retrieve Historical Monthly Price Data:

#### 1. Yahoo Finance:

- Go to [Yahoo Finance](#).
- Search for each ticker symbol (e.g., CHN.AX for Chalice Mining).
- Click on the stock and navigate to the "Historical Data" tab.
- Set the date range to the last five years and select "Monthly" frequency.
- Download the data as a CSV file for further analysis.

#### 2. Market Index:

- Visit [Market Index](#).
- Search for each company by name or ticker.
- Access the historical data section for monthly prices.


#### 3. TradingView:

- Go to [TradingView](#).
- Search for each stock by ticker.
- Use the chart settings to display historical data and select the monthly timeframe.
- Extract the data directly from the chart if needed.

Appendix D. Multi-Step Reasoning Task 9 Demonstration

The First ChatGPT-4o Trial	The Second Trial with Instruction	The Third Fresh Trial
<div><p>To calculate the annual rate of return for investors who bought Google shares at the IPO and held them until August 2018, we can use the formula for the compound annual growth rate (CAGR):</p><math display="block">CAGR = \left(\frac{FV}{PV}\right)^{\frac{1}{n}} - 1</math><p>where:</p><ul style="list-style-type: none"><li><math>FV</math> is the final value of the investment (\$1,084).</li><li><math>PV</math> is the initial value of the investment (\$85).</li><li><math>n</math> is the number of years the investment was held (2018 - 2004 = 14 years).</li></ul><p>Plugging the values into the formula:</p><math display="block">CAGR = \left(\frac{1084}{85}\right)^{\frac{1}{14}} - 1</math><p>Let's calculate it step by step:</p><ol style="list-style-type: none"><li>Calculate the ratio <math>\frac{FV}{PV}</math>: <math>\frac{1084}{85} \approx 12.7529</math></li><li>Calculate the exponent <math>\frac{1}{14}</math>: <math>12.7529^{\frac{1}{14}} \approx 1.2273</math></li><li>Subtract 1 from the result: <math>1.2273 - 1 = 0.2273</math></li><li>Convert to a percentage: <math>0.2273 \times 100 \approx 22.73\%</math></li></ol><p>So, the annual rate of return that investors earned from August 2004 to August 2018 is approximately <b>22.73%</b> per year.</p></div>	<div><p>what is 12.7529 ^ (1/14)?</p><p>The value of <math>12.7529^{\frac{1}{14}}</math> is approximately 1.1994.</p><p>To find the annual rate of return:</p><math display="block">CAGR = 1.1994 - 1 = 0.1994</math><p>Expressed as a percentage, this is:</p><math display="block">0.1994 \times 100 \approx 19.94\%</math><p>Therefore, the annual rate of return that investors who bought Google shares at the IPO and held them until August 2018 earned is approximately <b>19.94%</b>.</p><pre>python # Calculate the compound annual growth rate (CAGR) initial_value = 85 final_value = 1084 years = 2018 - 2004  # Calculate the growth factor growth_factor = final_value / initial_value  # Calculate the CAGR cagr = growth_factor ** (1 / years)  cagr</pre><p>Result</p><p>1.1994224359272951</p></div>	<div><p>The annual rate of return calculation using the CAGR formula can be demonstrated clearly.</p><p><b>Manual Calculation of CAGR:</b></p><ol style="list-style-type: none"><li><b>Define Initial and Final Prices:</b><ul style="list-style-type: none"><li>Initial Price (2004): \$85</li><li>Final Price (2018): \$1084</li></ul></li><li><b>Number of Years:</b><ul style="list-style-type: none"><li>Years = 2018 - 2004 = 14</li></ul></li><li><b>CAGR Formula:</b><ul style="list-style-type: none"><li><math display="block">CAGR = \left(\frac{\text{Final Price}}{\text{Initial Price}}\right)^{\frac{1}{\text{Years}}} - 1</math></li></ul></li></ol><p><b>Step-by-Step Calculation:</b></p><math display="block">CAGR = \left(\frac{1084}{85}\right)^{\frac{1}{14}} - 1</math><math display="block">CAGR = (12.75294118)^{\frac{1}{14}} - 1</math><math display="block">CAGR = (1.23520547) - 1</math><math display="block">CAGR = 0.1235 \text{ or } 12.35\%</math><p>Therefore, the annual rate of return that investors who bought Google shares at the IPO and held them until August 2018 earned is approximately 12.35%.</p><p>If you need further assistance or validation, please let me know!</p></div>

Appendix E. Multi-Step Reasoning Task 12 Demonstration

The First Trial	The Second Trial with Instruction (Correct Result)	The Third Fresh Trial (Wrong Result)
<div><div></div><div>To calculate the internal rate of return (IRR) for an investment, you need to find the discount rate that makes the net present value (NPV) of the investment equal to zero. The IRR is the rate <math>r</math> that satisfies the following equation:</div><div><math display="block">0 = -C_0 + \frac{C_1}{(1+r)^1} + \frac{C_2}{(1+r)^2} + \frac{C_3}{(1+r)^3} + \frac{C_4}{(1+r)^4} + \frac{C_5}{(1+r)^5}</math></div><div>where:</div><div><ul style="list-style-type: none"><li><math>C_0</math> is the initial investment (cost), which is \$800.</li><li><math>C_1, C_2, C_3, C_4, C_5</math> are the cash flows in years 1 to 5, which are \$300, \$200, \$150, \$122, and \$133 respectively.</li></ul></div><div>The equation to solve is:</div><div><math display="block">0 = -800 + \frac{300}{(1+r)^1} + \frac{200}{(1+r)^2} + \frac{150}{(1+r)^3} + \frac{122}{(1+r)^4} + \frac{133}{(1+r)^5}</math></div><div>Finding the IRR involves solving for <math>r</math>, which typically requires iterative methods or financial calculators. However, I will use numerical methods to approximate the IRR.</div><div>Let's solve this using a numerical method:</div><div><math display="block">0 = -800 + \frac{300}{(1+r)^1} + \frac{200}{(1+r)^2} + \frac{150}{(1+r)^3} + \frac{122}{(1+r)^4} + \frac{133}{(1+r)^5}</math></div><div>Calculation using a numerical method (trial and error or financial calculator):</div><div><div>1. Initial Guess: Let's start with an initial guess for <math>r</math>, say 10% (0.10):</div><div><math display="block">\text{NPV} = -800 + \frac{300}{1.10} + \frac{200}{1.10^2} + \frac{150}{1.10^3} + \frac{122}{1.10^4} + \frac{133}{1.10^5}</math><math display="block">\text{NPV} = -800 + 272.73 + 165.29 + 112.88 + 82.95 + 84.10</math><math display="block">\text{NPV} = -800 + 718.95</math><math display="block">\text{NPV} = -81.05</math></div></div><div>Since the NPV is negative, we need a lower rate. Let's try 15% (0.15):</div></div>	<div><div>6. Guess <math>r = 5.1\%</math> (0.051):</div><div><math display="block">\text{NPV} = -800 + \frac{300}{(1+0.051)^1} + \frac{200}{(1+0.051)^2} + \frac{150}{(1+0.051)^3} + \frac{122}{(1+0.051)^4} + \frac{133}{(1+0.051)^5}</math><math display="block">\text{NPV} = -800 + \frac{300}{1.051} + \frac{200}{1.104601} + \frac{150}{1.160951301} + \frac{122}{1.220158897} + \frac{133}{1.282309989}</math><math display="block">\text{NPV} = -800 + 285.44 + 181.11 + 129.23 + 100.03 + 103.74</math><math display="block">\text{NPV} = -800 + 799.55 = -0.45</math></div><div>Since the NPV is negative but very close to zero, we need a slightly lower rate.</div><div>7. Guess <math>r = 5.05\%</math> (0.0505):</div><div><math display="block">\text{NPV} = -800 + \frac{300}{(1+0.0505)^1} + \frac{200}{(1+0.0505)^2} + \frac{150}{(1+0.0505)^3} + \frac{122}{(1+0.0505)^4} + \frac{133}{(1+0.0505)^5}</math><math display="block">\text{NPV} = -800 + \frac{300}{1.0505} + \frac{200}{1.1036025} + \frac{150}{1.1578470625} + \frac{122}{1.21561084490625} + \frac{133}{1.27705629910625}</math><math display="block">\text{NPV} = -800 + 285.56 + 181.23 + 129.35 + 100.16 + 103.92</math><math display="block">\text{NPV} = -800 + 800.22 = 0.22</math></div><div>Since the NPV is slightly positive, we need a slightly higher rate, but we are very close. Let's try 5.075%.</div><div>8. Guess <math>r = 5.075\%</math> (0.05075):</div><div><math display="block">\text{NPV} = -800 + \frac{300}{(1+0.05075)^1} + \frac{200}{(1+0.05075)^2} + \frac{150}{(1+0.05075)^3} + \frac{122}{(1+0.05075)^4} + \frac{133}{(1+0.05075)^5}</math></div></div>	<div>Calculation of IRR:</div> <div><div>1. Define Cash Flows:</div><div><ul style="list-style-type: none"><li>Initial Investment: −\$800</li><li>Cash Flows: \$300, \$200, \$150, \$122, \$133 in years 1 to 5.</li></ul></div><div>2. Use the IRR formula:</div><div><ul style="list-style-type: none"><li><math>\sum \left( \frac{C_t}{(1+IRR)^t} \right) = 0</math></li></ul></div></div> <div>Python Code:</div> <div><pre>python import numpy as np  # Define the cash flows cash_flows = [-800, 300, 200, 150, 122, 133]  # Calculate the IRR irr = np.irr(cash_flows) irr_percentage = irr * 100 # Convert to percentage  print(f"The IRR for the investment is {irr_percentage:.2f}%")</pre></div> <div>Final IRR Value:</div> <div>Based on the provided cash flows, the IRR can be calculated as follows:</div> <div><div>IRR ≈ 12.36%</div><div>This is the approximate Internal Rate of Return for the given investment. If you need further assistance or validation, please let me know! [~]</div></div>

## Appendix F. 10-Stock Selection by ChatGPT-4

I would like to invest in Australian market, can you please give me a list of 10 best candidate stocks in Australia listed on the ASX exchange for the past three months from January till 14th of May 2024, and please also show me the criteria you have used to make the decision.

### Criteria Used:

1. **Market Capitalization:** Only included stocks with market caps above A\$500 million to exclude highly volatile penny stocks.
2. **Revenue Growth:** Focused on companies showing significant revenue growth, indicating expanding operations and market share.
3. **Profit Margins:** Considered companies with solid profit margins to ensure they are financially healthy.
4. **Historical Performance:** Analyzed the performance over the past year and five years to identify consistent growth patterns.
5. **Industry Leadership:** Selected companies that are leaders or strong contenders in their respective industries.

These stocks have been chosen based on their strong fundamentals and recent performance, but always conduct your own research or consult with a financial advisor to ensure they fit your investment strategy.



Searched 5 sites

Based on recent performance from January to May 2024, here are ten strong candidate stocks listed on the ASX (Australian Securities Exchange) that you might consider:

1. **Chalice Mining Limited (CHN)** - Specializes in mineral exploration, focusing on gold, copper, and nickel. It has shown a strong year-to-date performance.
2. **Vulcan Energy Resources Ltd (VUL)** - Engages in lithium exploration and geothermal energy. It has experienced significant growth, particularly in the development of its Zero Carbon Lithium project.
3. **Fineos Corporation Holdings Plc (FCL)** - Develops enterprise software for the insurance industry, providing solid growth figures.
4. **Southern Cross Gold Ltd (SXG)** - Focuses on gold and other precious metals exploration, showing substantial performance gains this year.
5. **Liontown Resources Ltd (LTR)** - Engages in the exploration of lithium and other minerals. It has demonstrated robust growth due to increased demand for lithium.
6. **Neuren Pharmaceuticals Ltd (NEU)** - Develops treatments for neurological disorders, achieving impressive returns due to successful drug trials and approvals.
7. **WiseTech Global (WTC)** - Provides logistics solutions, benefiting from the rising demand for global supply chain technologies.
8. **Aristocrat Leisure (ALL)** - A gaming technology company with strong performance in its gaming and mobile games segments.
9. **NextDC (NXT)** - Australia's largest data centre operator, seeing growth due to expanding data centre operations and partnerships.
10. **Pro Medicus Ltd (PME)** - A healthcare technology company specializing in imaging software, showing strong returns from its expanding international market ([Livewire Markets](#)) ([finder.com.au](#)) ([finder.com.au](#)) (IG).



## Appendix G. Complex Reasoning Task 1 Demonstration: Technical Analyses and Indicators for Chalice Mining Limited (CHN)

### Human analyst result—LSEG Workspace



Red is MACD line and blue is Signal line. Source: LSEG Refinitiv Workspace

### Recommendation

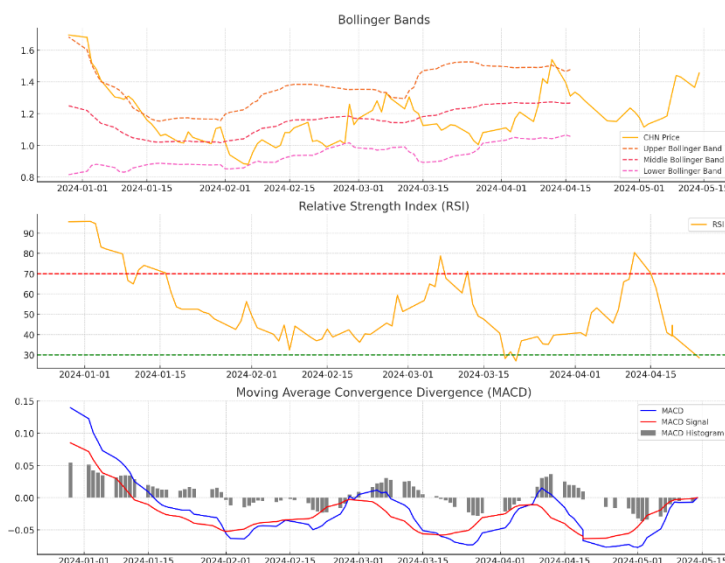
Bollinger Bands: price is close to the upper band and a bullish reversal has recently occurred, suggesting a “hold” at this stage.

RSI: it is approaching the upper limit “70”, may indicate “overbought” or “sell” situation.

MACD: MACD is above signal line, indicating “bullish” signal.

### ChatGPT-4o result

#### Technical Analysis for CHN



### Stock Recommendation Explanation

CHN Hold/Sell Overbought. MACD bullish. Price above upper Bollinger Band.

Appendix H. Complex Reasoning Task 2 Demonstration: Portfolio Construction

Global Minimum Variance Portfolio

Human analyst result

Stata:

```
. *global minimum variance portfolio*

. gmvport r_chn r_vul r_fcl r_sxg r_ltr r_neu r_wtc r_all r_nxt r_pme , noshort
Number of observations used to calculate expected returns and var-cov matrix : 93
The weight vector of the Global Minimum Variance Portfolio (NOT Allow Short Sales) is:
      Weights
r_chn      0
r_vul    .01650592
r_fcl    .05861241
r_sxg    .02194597
r_ltr    .00994657
r_neu    .08676104
r_wtc    .05172703
r_all    .5651326
r_nxt    .1346226
r_pme    .05474586
The return of the Global Minimum Variance Portfolio is: .00077709
The standard deviation (risk) of the Global Minimum Variance Portfolio is: .01055798
```

Excel:

Global Minimum Variance Portfolio CHN	VUL	FCL	SXG	LTR	NEU	WTC	ALL	NXT	PME	
Weight	0	0.016497079	0.058602942	0.021945695	0.009993004	0.086923706	0.051480691	0.565072901	0.134590389	0.054893592
Squared weight	0	0.000272154	0.003434305	0.000481614	9.98601E-05	0.007555731	0.002650262	0.319307384	0.018114573	0.003013306
Portfolio Return	0.00077348									
Portfolio Variance	0.00011147									
Portfolio Standard Deviation	0.01055798									
Risk-free Rate	0.00017									
Sharpe Ratio	0.05715859									

ChatGPT-4o result

Global Minimum Variance Portfolio

Optimal Weights:

Stock	Weight
CHN	0.00%
VUL	4.83%
FCL	10.94%
SXG	4.94%
LTR	4.69%
NEU	14.50%
WTC	14.97%
ALL	16.20%
NXT	13.83%
PME	15.11%

Global Minimum Variance Portfolio Return and Standard Deviation:

- Return: 0.1726%
- Standard Deviation: 1.2541%

## Optimal Risky Portfolio

## Human analyst result

## Stata:

```
. ovport r_chn r_vul r_fcl r_sxg r_ltr r_neu r_wtc r_all r_nxt r_pme , rfrate(0.00017) noshort
Number of observations used to calculate expected returns and var-cov matrix : 93
The weight vector of the Tangent Portfolio with a risk-free rate of .00017 (NOT Allow Short Sales) is:

Weights
r_chn      0
r_vul    .06273422
r_fcl      0
r_sxg    .18148295
r_ltr      0
r_neu      0
r_wtc    .25450635
r_all      0
r_nxt    .3488319
r_pme    .15244458
The return of the Tangent Portfolio is: .00447312
The standard deviation (risk) of the Tangent Portfolio is: .01712425
```

## Excel:

Optimal Risky Portfolio	CHN	VUL	FCL	SXG	LTR	NEU	WTC	ALL	NXT	PME
Weight	0	0.063442466		0	0.183980533	0	0	0.253877687	0	0.348578314
Squared weight	0	0.004024947		0	0.033648837	0	0	0.06445388	0	0.121506841
Portfolio Return	0.00448193									
Portfolio Variance	0.00029651									
Portfolio Standard Deviation	0.01721945									
Risk-free Rate	0.00017									
Sharpe Ratio	0.25041038									

## ChatGPT-4o result

## Optimal Risky Portfolio

## Optimal Weights:

Stock	Weight
CHN	0.00%
VUL	6.37%
FCL	0.00%
SXG	18.40%
LTR	0.00%
NEU	0.00%
WTC	25.44%
ALL	0.00%
NXT	34.85%
PME	14.94%

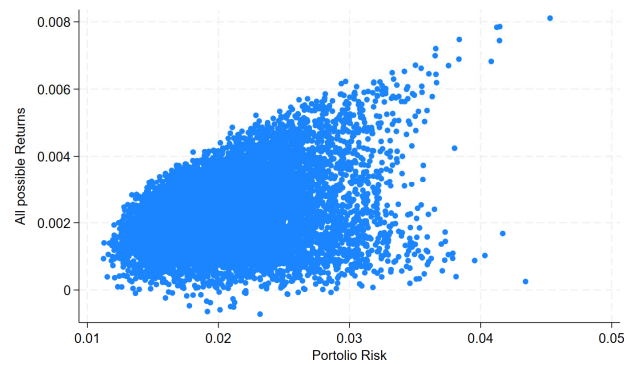
## Optimal Risky Portfolio Return and Standard Deviation:

- Return: 0.4483% per day
- Standard Deviation: 1.7224% per day

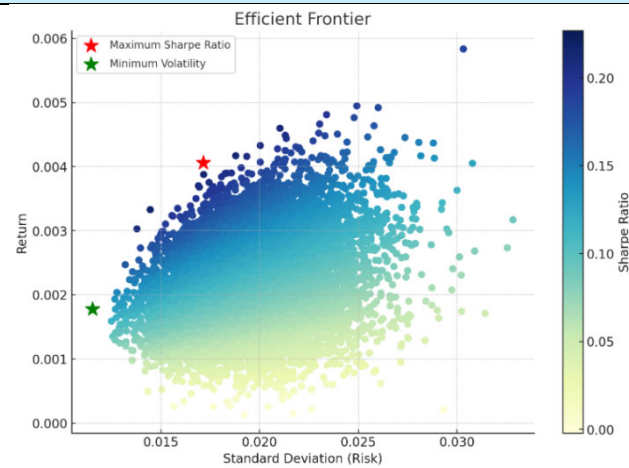
## Efficient Frontier

## Human analyst result

Stata:



## ChatGPT-4o result



## Appendix I. Complex Reasoning Task 3 Demonstration: Capital Budgeting

## Human analyst result

Cash Flows	0	1	2	3	4	5	6
Revenues		\$480,000.00	\$480,000.00	\$480,000.00	\$480,000.00	\$480,000.00	\$480,000.00
- Operating Costs		\$100,000.00	\$100,000.00	\$100,000.00	\$100,000.00	\$100,000.00	\$100,000.00
- Depreciation		\$420,000.00	\$252,000.00	\$151,200.00	\$90,720.00	\$54,432.00	\$32,659.20
Taxable income		-\$40,000.00	\$128,000.00	\$228,800.00	\$289,280.00	\$325,568.00	\$347,340.80
- Taxes (32%)		-\$12,800.00	\$40,960.00	\$73,216.00	\$92,569.60	\$104,181.76	\$111,149.06
Earnings		-\$27,200.00	\$87,040.00	\$155,584.00	\$196,710.40	\$221,386.24	\$236,191.74
Operating CFs (Earn + Depr)		\$392,800.00	\$339,040.00	\$306,784.00	\$287,430.40	\$275,818.24	\$268,850.94
- Sublease lost		\$80,000.00	\$80,000.00	\$80,000.00	\$80,000.00	\$80,000.00	\$80,000.00
- Cannibalisation		\$25,000.00	\$25,000.00	\$25,000.00	\$25,000.00	\$25,000.00	\$25,000.00
Net Cash Flow		\$287,800.00	\$234,040.00	\$201,784.00	\$182,430.40	\$170,818.24	\$163,850.94
Answer to 10. NPV							
@ 12% Discount Rate							
	0	1	2	3	4	5	6
Initial cash flow (excluding change in NWC)	-\$1,050,000.00						
Net Cash Flow		\$287,800.00	\$234,040.00	\$201,784.00	\$182,430.40	\$170,818.24	\$163,850.94
Terminal Value							\$833,333.33
Total Project Value	-\$1,050,000.00	\$287,800.00	\$234,040.00	\$201,784.00	\$182,430.40	\$170,818.24	\$997,184.28
NPV		\$255,234.67					
@ 14% Discount Rate							
	0	1	2	3	4	5	6
Initial cash flow (excluding change in NWC)	-\$1,050,000.00						
Net Cash Flow		\$287,800.00	\$234,040.00	\$201,784.00	\$182,430.40	\$170,818.24	\$163,850.94
Terminal Value							\$714,285.71
Total Project Value	-\$1,050,000.00	\$287,800.00	\$234,040.00	\$201,784.00	\$182,430.40	\$170,818.24	\$878,136.66
NPV		\$115,539.10					

## ChatGPT-4o result

## 1. Determine the initial investment:

- Purchase of 100 used cars: 100 cars \* \$9,500 per car = \$950,000
- Installation of LoJack systems: 100 cars \* \$1,000 per car = \$100,000
- Total initial investment = \$950,000 + \$100,000 = \$1,050,000

## 2. Calculate annual revenues and costs:

- Annual revenue per car: \$4,800
- Total annual revenue: 100 cars \* \$4,800 = \$480,000
- Annual operating cost per car: \$1,000
- Total annual operating cost: 100 cars \* \$1,000 = \$100,000
- Opportunity cost of subleasing the lot: \$80,000
- Lot maintenance cost: \$20,000 (already accounted for in both scenarios)
- Net cash flow from regular rental car business: -\$25,000

## 3. Calculate depreciation:

- Using the diminishing value method, the formula for annual depreciation is:  

$$\text{Depreciation} = \text{Cost} \times \left( \frac{2}{\text{Useful Life}} \right)$$
- First-year depreciation:  $\text{Depreciation}_{\text{Year 1}} = \$1,050,000 \times \left( \frac{2}{5} \right) = \$420,000$
- Subsequent years: Calculate the depreciation for each year (Year 2 to Year 5) based on the book value at the start of each year.

## 4. Calculate net cash flows:

- Net revenue before tax: Total revenue – Operating costs – Opportunity cost – Lost cash flow
- Calculate taxable income by subtracting depreciation from net revenue before tax.
- Calculate taxes:  $\text{Tax} = \text{Taxable income} \times 32\%$
- Calculate net cash flow after tax: Net revenue before tax – Taxes + Depreciation
- Include change in net working capital in the first year.

## 5. Calculate terminal value and net working capital recovery:

- Terminal value after 5 years: Perpetual net cash flow of \$100,000 per year discounted at the adjusted discount rate.
- Recover net working capital in the terminal year.

## 6. Calculate NPV:

- Use the discount rates of 12% and 14% to calculate the present value of cash flows and terminal value.

## Capital Budgeting Analysis

## Initial Investment

- Purchase of 100 used cars: \$950,000
- Installation of LoJack systems: \$100,000
- Total initial investment: \$1,050,000
- Net Working Capital: \$50,000 (initially subtracted from Year 1 cash flows)

## Annual Revenues and Costs

- Annual revenue: \$480,000
- Annual operating costs: \$100,000
- Opportunity cost of subleasing lot: \$80,000
- Net cash flow impact on regular rental car business: -\$25,000
- Net revenue before tax: \$275,000

## Depreciation (Diminishing Value Method)

- Year 1: \$420,000
- Year 2: \$252,000
- Year 3: \$151,200
- Year 4: \$90,720
- Year 5: \$54,432

## Net Cash Flows (After Tax)

- Year 1: \$112,000
- Year 2: \$275,360
- Year 3: \$344,408
- Year 4: \$368,692
- Year 5: \$377,187

## Terminal Value (Perpetuity starting from Year 6)

- Terminal Value:  $\$100,000 / 0.12 = \$833,333.33$

## Net Present Value (NPV)

## Using 12% Discount Rate

- NPV = \$1,071,417.83

## Using 14% Discount Rate

- NPV = \$966,137.31

## Conclusion

- At a 12% discount rate, the NPV of the project is \$1,071,417.83, indicating that the project is profitable.
- At a 14% discount rate, the NPV of the project is \$966,137.31, still indicating profitability, but with a reduced margin due to the higher risk reflected in the higher discount rate.

These NPVs suggest that ACE should consider undertaking the project, especially if the risk-adjusted NPV (at 14%) is still significantly positive. [\[3\]](#)

## Appendix J. Complex Reasoning Task 4 Demonstration: Financial Statement Analysis

## Human analyst result

Answer:

$$\begin{aligned} \text{a. ROE} &= \frac{\text{Net profit}}{\text{Equity}} = \frac{\text{Net profit}}{\text{Pretax profit}} \times \frac{\text{Pretax profit}}{\text{EBIT}} \times \frac{\text{EBIT}}{\text{Sales}} \times \frac{\text{Sales}}{\text{Assets}} \times \frac{\text{Assets}}{\text{Equity}} \\ &= \text{Tax burden} \times \text{Interest burden} \times \text{Profit margin} \times \text{Asset turnover} \times \text{Leverage} \\ \text{Tax burden} &= \frac{\text{Net profit}}{\text{Pretax Profit}} = \frac{\$510}{\$805} = .6335 \\ \text{Interest burden} &= \frac{\text{Pretax profit}}{\text{EBIT}} = \frac{\$805}{\$830} = .9699 \\ \text{Profit margin} &= \frac{\text{EBIT}}{\text{Sales}} = \frac{\$830}{\$5,140} = .1615 \\ \text{Asset turnover} &= \frac{\text{Sales}}{\text{Average Total Assets}} = \frac{\$5,140}{(\$3,100 + \$2,950)/2} = 1.6992 \\ \text{Leverage} &= \frac{\text{Average Total Assets}}{\text{Average Total Equity}} = \frac{(\$3,100 + \$2,950)/2}{(\$2,200 + \$2,100)/2} = 1.4070 \\ \text{b. ROE} &= .6335 \times .9699 \times .1615 \times 1.6992 \times 1.4070 = .2372 = 23.72\% \\ \text{c. } g &= \text{ROE} \times \text{plowback} = .2372 \times \frac{\$1.96 - \$ .60}{\$1.96} = .1646 = 16.46\% \end{aligned}$$

## ChatGPT-4o result

## Excel template created by GPT-4o:

	2021	2022
Revenue	4750	5140
Cost of goods sold	2400	2540
Selling, general, and administrative	1400	1550
Depreciation	180	210
Goodwill amortization	10	10
Operating income	760	830
Interest expense	20	25
Income before taxes	740	805
Income taxes	265	295
Net income	475	510
Earnings per share	1.79	1.96
Average shares outstanding (millions)	265	260
Cash	400	400
Accounts receivable	680	700
Inventories	570	600
Net property, plant, and equipment	800	820
Intangibles	500	530
Total assets	2950	3100
Current liabilities	550	600
Long-term debt	300	300
Total liabilities	850	900
Stockholders' equity	2100	2200
Total liabilities and equity	2950	3100
Book value per share	8.00	8.46
Annual dividend per share	0.55	0.58

## ChatGPT-4o's calculations:

## a. Components of the DuPont Formula for 2022

1. Net Profit Margin:  

$$\text{Net Profit Margin} = \frac{\text{Net Income}}{\text{Revenue}} = \frac{510}{5140} \approx 0.099$$
2. Asset Turnover:  

$$\text{Asset Turnover} = \frac{\text{Revenue}}{\text{Total Assets}} = \frac{5140}{3100} \approx 1.658$$
3. Equity Multiplier:  

$$\text{Equity Multiplier} = \frac{\text{Total Assets}}{\text{Stockholders' Equity}} = \frac{3100}{2200} \approx 1.409$$

## b. ROE for 2022

Using the DuPont formula:

$$\text{ROE} = \text{Net Profit Margin} \times \text{Asset Turnover} \times \text{Equity Multiplier}$$

$$\text{ROE} = 0.099 \times 1.658 \times 1.409 \approx 0.232 \text{ or } 23.2\%$$

## c. Sustainable Growth Rate for 2022

1. Dividend Payout Ratio:  

$$\text{Dividend Payout Ratio} = \frac{\text{Annual Dividend per Share}}{\text{Earnings per Share}} = \frac{0.58}{1.96} \approx 0.296$$
2. Plowback Ratio:  

$$\text{Plowback Ratio} = 1 - \text{Dividend Payout Ratio} = 1 - 0.296 \approx 0.704$$
3. Sustainable Growth Rate (SGR):  

$$\text{SGR} = \text{ROE} \times \text{Plowback Ratio} = 0.232 \times 0.704 \approx 0.163 \text{ or } 16.3\%$$

So, the sustainable growth rate for 2022 is approximately 16.3%. [5]

## Appendix K. Complex Reasoning Task 5 Demonstration: Financial Statement Analysis

## Human analyst result

Answer:

- a. WhiteBrush has had higher sales and earnings growth (per share) than ProtectWhite. Margins are also higher. But this does not necessarily mean that WhiteBrush is a better investment. ProtectWhite has a higher ROE, which has been stable, while WhiteBrush's ROE has been declining. We can use DuPont analysis to identify the source of the difference in ROE:

Component	Definition	WhiteBrush	ProtectWhite
Tax burden	Net profit/Pretax profit	67.44%	65.99%
Interest burden	Pretax profit/EBIT	1.00	0.9545
Profit margin	EBIT/Sales	8.51%	6.46%
Asset turnover	Sales/Average Assets	1.8259	3.6286
Leverage	Average Assets/Average Equity	1.5071	1.5386
ROE	Net profit/Average Equity	15.8%	22.7%

While tax burden, interest burden, and leverage are similar, profit margin and asset turnover differ. Although ProtectWhite has a lower profit margin, it has far higher asset turnover.

Sustainable Growth = ROE × Plowback Ratio

	ROE	Plowback ratio	Sustainable growth rate	Ludlow's estimate
WhiteBrush	15.8%	1.00	15.8%	30.0%
ProtectWhite	22.7%	.344	7.8%	10.0%

Jennifer has overestimated the sustainable growth rate for each company. WhiteBrush has little ability to increase its sustainable growth because plowback already equals 100%. ProtectWhite could increase its sustainable growth by increasing its plowback ratio.

- b. WhiteBrush's recent EPS growth has been achieved by increasing book value per share, not by achieving greater profits per dollar of equity. Since EPS is equal to (Book value per share × ROE), a firm can increase EPS even if ROE is declining; this is the case for WhiteBrush. WhiteBrush's book value per share has more than doubled in the last two years.

Book value per share can increase either by retaining earnings or by issuing new stock at a market price greater than book value. WhiteBrush has been retaining all earnings, but the increase in the number of outstanding shares indicates that it has also issued a substantial amount of stock.

## ChatGPT-4o result

ChatGPT-4o’s Parts a—i calculations:

WhiteBrush:		ProtectBrush:	
Calculations for WhiteBrush:	Equity Multiplier (2020, 2021, 2022): $\text{Equity Multiplier} = \frac{\text{Total Assets}}{\text{Shareholder's Equity}}$ <ul style="list-style-type: none"><li>2020: <math>\frac{2660}{600} \approx 4.43</math></li><li>2021: <math>\frac{3300}{980} \approx 3.37</math></li><li>2022: <math>\frac{5470}{1750} \approx 3.12</math></li></ul> ROE (2020, 2021, 2022): $\text{ROE} = \text{Net Profit Margin} \times \text{Asset Turnover} \times \text{Equity Multiplier}$ <ul style="list-style-type: none"><li>2020: <math>5.46\% \times 1.31 \times 4.43 \approx 31.77\%</math></li><li>2021: <math>5.37\% \times 1.64 \times 3.37 \approx 29.70\%</math></li><li>2022: <math>5.73\% \times 1.42 \times 3.12 \approx 25.38\%</math></li></ul>	Calculations for ProtectBrush:	Equity Multiplier (2020, 2021, 2022): <ul style="list-style-type: none"><li>2020: <math>\frac{33700}{18800} \approx 1.79</math></li><li>2021: <math>\frac{31200}{10200} \approx 3.06</math></li><li>2022: <math>\frac{33000}{10000} \approx 3.30</math></li></ul> ROE (2020, 2021, 2022): <ul style="list-style-type: none"><li>2020: <math>3.94\% \times 3.09 \times 1.79 \approx 21.75\%</math></li><li>2021: <math>3.94\% \times 3.54 \times 3.06 \approx 42.65\%</math></li><li>2022: <math>3.73\% \times 3.61 \times 3.30 \approx 44.51\%</math></li></ul>
Net Profit Margin (2020, 2021, 2022): $\text{Net Profit Margin} = \frac{\text{Net Income}}{\text{Revenue}}$ <ul style="list-style-type: none"><li>2020: <math>\frac{190}{3480} \approx 5.46\%</math></li><li>2021: <math>\frac{290}{5400} \approx 5.37\%</math></li><li>2022: <math>\frac{445}{7760} \approx 5.73\%</math></li></ul>		Net Profit Margin (2020, 2021, 2022): <ul style="list-style-type: none"><li>2020: <math>\frac{4100}{104000} \approx 3.94\%</math></li><li>2021: <math>\frac{4350}{110400} \approx 3.94\%</math></li><li>2022: <math>\frac{4450}{119200} \approx 3.73\%</math></li></ul>	
Asset Turnover (2020, 2021, 2022): $\text{Asset Turnover} = \frac{\text{Revenue}}{\text{Total Assets}}$ <ul style="list-style-type: none"><li>2020: <math>\frac{3480}{2660} \approx 1.31</math></li><li>2021: <math>\frac{5400}{3300} \approx 1.64</math></li><li>2022: <math>\frac{7760}{5470} \approx 1.42</math></li></ul>		Asset Turnover (2020, 2021, 2022): <ul style="list-style-type: none"><li>2020: <math>\frac{104000}{33700} \approx 3.09</math></li><li>2021: <math>\frac{110400}{31200} \approx 3.54</math></li><li>2022: <math>\frac{119200}{33000} \approx 3.61</math></li></ul>	

ChatGPT-4o’s Parts a—ii calculations:

WhiteBrush:	ProtectBrush:	Analysis:
Calculations for WhiteBrush:	Calculations for ProtectBrush:	
Dividend Payout Ratio (2020, 2021, 2022): <ul style="list-style-type: none"><li>2020: <math>\frac{0.30 \times 317}{190} \approx 0.50</math> or 50%</li><li>2021: <math>\frac{0.30 \times 346}{290} \approx 0.36</math> or 36%</li><li>2022: <math>\frac{0.30 \times 376}{445} \approx 0.25</math> or 25%</li></ul>	Dividend Payout Ratio (2020, 2021, 2022): <ul style="list-style-type: none"><li>2020: <math>\frac{3 \times 1850}{4100} \approx 1.35</math> or 135%</li><li>2021: <math>\frac{3 \times 1850}{4350} \approx 1.28</math> or 128%</li><li>2022: <math>\frac{3 \times 1850}{4450} \approx 1.25</math> or 125%</li></ul>	<ul style="list-style-type: none"><li>Profitability (ROE): ProtectBrush has a higher ROE compared to WhiteBrush, especially in 2021 and 2022, suggesting better utilization of equity to generate profits.</li><li>Sustainable Growth Rate (SGR): WhiteBrush has a positive SGR, while ProtectBrush has a negative SGR due to its excessively high dividend payout, leading to negative plowback ratios.</li></ul>
Plowback Ratio (2020, 2021, 2022): <ul style="list-style-type: none"><li>2020: <math>1 - 0.50 = 0.50</math> or 50%</li><li>2021: <math>1 - 0.36 = 0.64</math> or 64%</li><li>2022: <math>1 - 0.25 = 0.75</math> or 75%</li></ul>	Since the dividend payout ratio exceeds 100%, ProtectBrush is paying out more than its earnings in dividends, resulting in a negative plowback ratio.  SGR (2020, 2021, 2022): <ul style="list-style-type: none"><li>2020: <math>21.75\% \times (-0.35) = -7.61\%</math></li><li>2021: <math>42.65\% \times (-0.28) = -11.94\%</math></li><li>2022: <math>44.51\% \times (-0.25) = -11.13\%</math></li></ul>	Jennifer's analysis is flawed because: <ul style="list-style-type: none"><li>WhiteBrush is not necessarily more profitable based on ROE. ProtectBrush has a higher ROE.</li><li>The claim of a higher sustainable growth rate for WhiteBrush is correct, but it stems from ProtectBrush's negative SGR rather than WhiteBrush's superiority.</li></ul>
SGR (2020, 2021, 2022): <ul style="list-style-type: none"><li>2020: <math>31.77\% \times 0.50 = 15.89\%</math></li><li>2021: <math>29.70\% \times 0.64 = 19.01\%</math></li><li>2022: <math>25.38\% \times 0.75 = 19.04\%</math></li></ul>		

ChatGPT-4o’s Part b explanations:

<p><b>Part b: Explaining EPS Growth Despite Declining ROE</b></p> <p>WhiteBrush has produced an average annual EPS growth rate of 40% over the last two years despite a declining ROE. Possible reasons include:</p> <ol style="list-style-type: none"><li><b>Revenue Growth:</b> WhiteBrush has shown substantial revenue growth (from \$3,480 in 2020 to \$7,760 in 2022), which has increased the overall earnings.</li><li><b>Cost Management:</b> The company has managed its costs effectively, maintaining a relatively stable profit margin.</li><li><b>Share Buybacks:</b> The average shares outstanding increased marginally, suggesting that the company might be using some of its earnings to repurchase shares, thereby boosting EPS.</li><li><b>Dividend Payout:</b> A lower dividend payout ratio in recent years indicates that more earnings are retained within the company, contributing to EPS growth.</li></ol> <p>These factors combined can lead to a higher EPS even if the ROE is declining.</p>
---

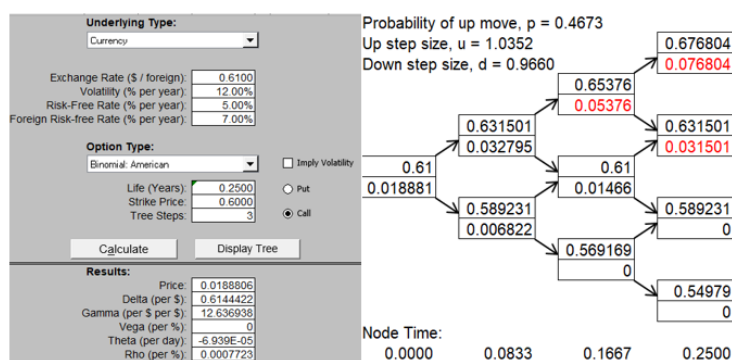


Category	December 2020	December 2021	December 2022	3-Year Average
Income statement				
Revenue	3480	5400	7760	
Cost of goods sold	2700	4270	6050	
Selling, general, and admin. expense	500	690	1000	
Depreciation and amortization	250	400	660	
Operating income (EBIT)	30	40	30	
Interest expense	250	400	660	
Income before taxes	20	40	35	
Income taxes	60	110	215	
Income after taxes	190	290	445	
Diluted EPS	0.6	0.84	1.18	
Average shares outstanding (000)	317	346	376	
Financial statistics				
COGS as % of sales	77.59	79.07	77.96	78.24
General & admin. as % of sales	14.37	12.78	13.32	13.16
Operating margin (%)	7.18	7.41	8.51	7.7
Pretax income/EBIT (%)	100	100	100.8	100.27
Tax rate (%)	24	27.5	32.58	28.69
Balance sheet				
Cash and cash equivalents	460	50	480	
Accounts receivable	540	720	950	
Inventories	760	830	1340	
Net property, plant, and equipment	300	1830	3450	
Total assets	2660	3300	5470	
Current liabilities	860	1110	1750	
Long-term debt	1200	1210	1970	
Total liabilities	2060	2320	3720	
Stockholders' equity	600	980	1750	
Total liabilities and equity	2660	3300	5470	
Market price per share	21	30	45	
Book value per share	1.89	2.84	4.66	
Annual dividend per share	0.3	0.3	0.3	

Category	December 2020	December 2021	December 2022	3-Year Average
Income statement				
Revenue	104000	110400	119200	
Cost of goods sold	72800	75000	79300	
Selling, general, and admin. expense	20300	22000	23900	
Depreciation and amortization	4200	4700	4700	
Operating income (EBIT)	6700	6900	7500	
Interest expense	500	550	550	
Income before taxes	6200	6550	6950	
Income taxes	2100	2200	2500	
Income after taxes	4100	4350	4450	
Diluted EPS	2.16	2.35	2.41	
Average shares outstanding (000)	1850	1850	1850	
Financial statistics				
COGS as % of sales	70	68	66.53	68.1
General & admin. as % of sales	19.52	20.64	20.05	20.08
Operating margin (%)	6.44	6.25	6.29	6.33
Pretax income/EBIT (%)	91.04	94.25	95.45	93.58
Tax rate (%)	34.43	33.59	34.01	34.01
Balance sheet				
Cash and cash equivalents	7900	3300	1000	
Accounts receivable	7500	3600	8000	
Inventories	1200	1800	1600	
Net property, plant, and equipment	12000	14500	17000	
Total assets	33700	31200	33000	
Current liabilities	13000	14000	16000	
Long-term debt	9500	7000	7000	
Total liabilities	22500	21000	23000	
Stockholders' equity	18800	10200	10000	
Total liabilities and equity	33700	31200	33000	
Market price per share	10	18	20	
Book value per share	10.16	5.51	5.41	
Annual dividend per share	3	3	3	

## Appendix L. Complex Reasoning Task 6 Demonstration: Option Pricing – Binomial Tree

### Human analyst result



At each node: upper value = underlying asset price;  
lower value = option price and values in red are a result of early exercise.

### ChatGPT-4o result



**ChatGPT-4o voice model:**

The black boxes contain the asset prices at each node.

The dashed boxes contain the option prices at each node.

Blue dashed lines indicate the option prices when holding the option is optimal.

Red dashed lines indicate the option prices when early exercise is optimal.

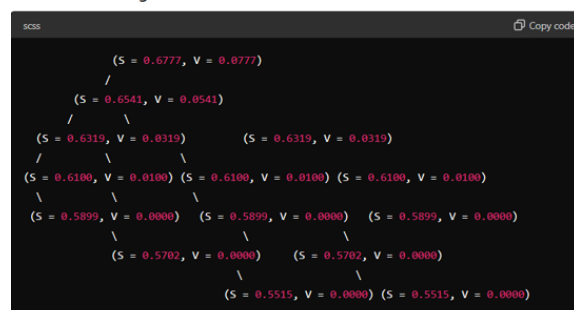
Up factor:  $u = e^{\sigma\sqrt{\Delta t}} = e^{0.12\sqrt{0.0833}} \approx 1.0342$

Down factor:  $d = \frac{1}{u} = \frac{1}{1.0342} \approx 0.967$

Risk-neutral probability:

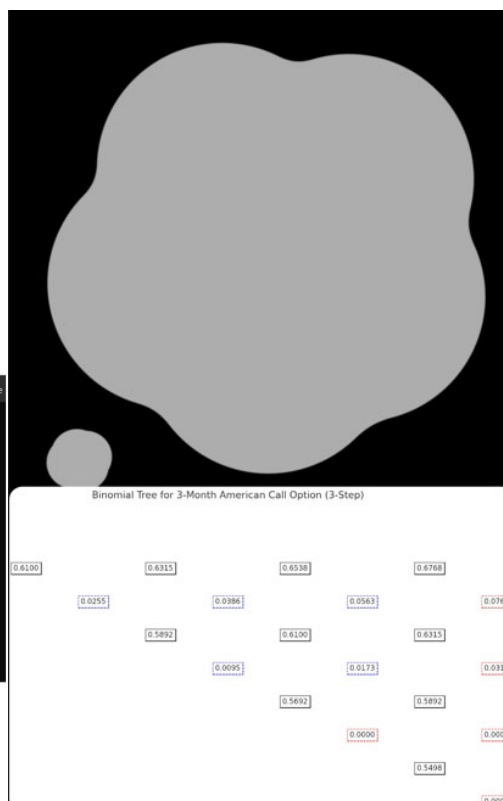
$$p = \frac{e^{(r_d - r_f)\Delta t} - d}{u - d} = \frac{e^{(0.05 - 0.07) \cdot 0.0833} - 0.967}{1.0342 - 0.967} \approx 0.4886$$

Binomial Tree Diagram



In this diagram, nodes with  $V$  in red indicate where early exercise is optimal. Given the current values and parameters, there are no nodes where early exercise would be optimal since the option values are non-negative and less than the intrinsic values at every point.

Therefore, early exercise is not optimal at any step.



## References

- Ahmed, S., M. M. Alshater, A. El Ammari, and H. Hammami. 2022. Artificial Intelligence and Machine Learning in Finance: A Bibliometric Review. *Research in International Business and Finance* 61: 101646. <https://doi.org/10.1016/j.ribaf.2022.101646>.
- Blesiada, Jamie. 2023. Expedia Group Gives Users the Opportunity to Test New Technology. *Tavel Weekly*. Available online: <https://www.travelweekly.com/Travel-News/Travel-Technology/Expedia-Group-gives-users-opportunity-test-new-technology> (accessed on 16 May 2024).
- Bodie, Z., A. Kane, and A. J. Marcus. 2022. *Essentials of Investments*, 12th ed. New York: McGraw Hill LLC.
- Bollen, J., H. Mao, and X. Zeng. 2011. Twitter Mood Predicts the Stock Market. *Journal of Computational Science* 2: 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>.
- Boukes, M., B. Van de Velde, T. Araujo, and R. Vliegthart. 2020. What's the Ttone? Easy doesn't Do It: Analyzing Performance and Agreement between Off-the-shelf Sentiment Analysis Tools. *Communication Methods and Measures* 14: 83–104.
- Brealey, R. A., S. C. Myers, F. Allen, and A. Edmans. 2022. *Principles of Corporate Finance*, 14th ed. New York: McGraw-Hill US.
- Burgess, N. 2021. Machine Earning—Algorithmic Trading Strategies for Superior Growth, Outperformance and Competitive Advantage. *International Journal of Artificial Intelligence and Machine Learning* 2: 38–60.
- Cao, L. 2022. AI in Finance: Challenges, Techniques, and Opportunities. *ACM Computing Surveys (CSUR)* 55: 1–38. <https://doi.org/10.1145/3502289>.
- Chaboud, A. P., B. Chiquoine, E. Hjalmarsson, and C. Vega. 2014. Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market. *The Journal of Finance* 69: 2045–84. <https://doi.org/10.1111/jofi.12186>.
- Cheng, L., X. Li, and L. Bing. 2023. Is GPT-4 a Good Data Analyst? *arXiv*, arXiv:2305.15038. <https://doi.org/10.48550/arXiv.2305.15038>.
- de Lange, P. E., B. Melsom, C. B. Vennerød, and S. Westgaard. 2022. Explainable AI for credit assessment in banks. *Journal of Risk and Financial Management* 15: 556.

- Demajo, L. M., V. Vella, and A. Dingli. 2020. Explainable AI for Interpretable Credit Scoring. *arXiv*, arXiv:2012.03749. <https://doi.org/10.5121/csit.2020.101516>.
- Dilmegani, C. 2024. ChatGPT Code Interpreter Plugin: Use Cases & Limitations in 2024. *AIMultiple Research*. Available online: <https://research.aimultiple.com/chatgpt-code-interpreter/> (accessed on 16 May 2024).
- Dowling, M., and B. Lucey. 2023. ChatGPT for (Finance) Research: The Bananarama Conjecture. *Finance Research Letters* 53: 103662. <https://doi.org/10.1016/j.frl.2023.103662>.
- Elad, B. 2024. OpenAI Statistics 2024 By Demographics, Products, Revenue and Growth. Available online: [https://www.enterpriseposttoday.com/stats/openai-statistics.html#google\\_vignette](https://www.enterpriseposttoday.com/stats/openai-statistics.html#google_vignette) (accessed on 11 June 2024).
- Farooq, A., and P. Chawla. 2021. Review of Data Science and AI in Finance. Paper presented at International Conference on Computing Sciences (ICCS), Phagwara, India, December 4–5. <https://doi.org/10.1109/ICCS54944.2021.00050>.
- Félix, L., R. Kräussl, and P. Stork. 2020. Implied Volatility Sentiment: A Tale of Two Tails. *Quantitative Finance* 20: 823–49. <https://doi.org/10.1080/14697688.2019.1696018>.
- Hansen, A. L., and S. Kazinnik. 2023. Can ChatGPT Decipher FedSpeak? Federal Reserve Bank of New York. United States of America. Available online: <https://policycommons.net/artifacts/5671671/can-chatgpt-decipher-fedspeak/6437313/> (accessed on 11 June 2024).
- Hartmann, J., M. Heitmann, C. Siebert, and C. Schamp. 2023. More Than a Feeling: Accuracy and Application of Sentiment Analysis. *International Journal of Research in Marketing* 40: 75–87. <https://doi.org/10.1016/j.ijresmar.2022.05.005>.
- Hull, J. C. 2015. *Options, Futures, and Other Derivatives, Global Edition*. London: Pearson Education.
- Jullum, M., A. Løland, R. B. Huseby, Ånonsen, G., and J. Lorentzen. 2020. Detecting Money Laundering Transactions with Machine Learning. *Journal of Money Laundering Control* 23: 173–86.
- Kelly, Jack. 2023. Goldman Sachs Predicts 200 Million Jobs will be Lost or Degraded by Artificial Intelligence. *Forbes*. Accessed online: <https://www.forbes.com/sites/jackkelly/2023/03/31/goldman-sachs-predicts-300-million-jobs-will-be-lost-or-degraded-by-artificial-intelligence/?sh=43cb004a782b> (accessed on 16 May 2024).
- Kocor, J., I. Cichecki, O. Kaszyca, M. Kochanek, D. Szydło, J. Baran, Julita Bielaniec, Marcin Gruza, Arkadiusz Janz, K. Kanclerz, and et al. 2023. ChatGPT: Jack of All Trades, Master of None. *Information Fusion* 99: 101861.
- Leippold, M. 2023a. Sentiment Spin: Attacking Financial Sentiment with GPT-3. *Finance Research Letters* 55: 103957. <https://doi.org/10.1016/j.frl.2023.103957>.
- Leippold, M. 2023b. Thus spoke GPT-3: Interviewing a large-language model on climate finance. *Finance Research Letters* 53: 103617.
- Lin, T. C. 2019. Artificial Intelligence, Finance, and the Law. *Fordham Law Review* 88: 531.
- Liu, L. X., S. Liu, and M. Sathye. 2021. Predicting Bank Failures: A Synthesis of Literature and Directions for Future Research. *Journal of Risk and Financial Management* 14: 474. <https://doi.org/10.3390/jrfm14100474>.
- Lopez-Lira, A., and Y. Tang. 2023. Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models. *arXiv*, arXiv:2304.07619. <https://doi.org/10.48550/arXiv.2304.07619>.
- Masson, D. J. 2018. 6 Steps to An Effective Financial Statement Analysis. *Association for Financial Professionals*. Available online: <https://www.afponline.org/training-resources/resources/articles/Details/6-steps-to-an-effective-financial-statement-analysis> (accessed on 16 May 2024).
- Roumeliotis, K. I., and N. D. Tselikas. 2023. Chatgpt and Open-AI models: A Preliminary Review. *Future Internet* 15: 192.
- Saunders, A., M. M. Cornett, and O. Erhemjams. 2021. *Financial Institutions Management: A Risk Management Approach*, 10th ed. New York: McGraw-Hill Education.
- Selke, M. J. G. 2013. *Rubric Assessment Goes to College: Objective, Comprehensive Evaluation of Student Work*. Lanham: R&L Education.
- Sokolov, I. 2019. Theory and Practice in Artificial Intelligence. *Вестник Российской академии наук* 89: 365–70.
- Son, G., H. Jung, M. Hahm, K. Na, and S. Jin. 2023. Beyond Classification: Financial Reasoning in State-of-the-Art Language Models. *arXiv*, arXiv:2305.01505.
- Stevens, D. D., and A. J. Levi. 2023. *Introduction to rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning*. Abingdon-on-Thames: Routledge.
- Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance* 63: 1437–67.
- Turing, A. M. 2012. Computing Machinery and Intelligence (1950). The Essential Turing: the Ideas That Gave Birth to the Computer Age 49: 433–64. Available online: <https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf> (accessed on 16 May 2024).
- Wahlen, J. M., S. P. Baginski, and M. T. Bradshaw. 2018. *Financial Reporting, Rincial Statement Analysis, and Valuation: A Strategic Perspective*. Boston: Cengage Learning.
- Wei, J., X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35: 24824–37.
- Wenzlaff, K., and S. Spaeth. 2022. Smarter than Humans? Validating how OpenAI's ChatGPT Model Explains Crowdfunding, Alternative Finance and Community Finance. *Validating How OpenAI's ChatGPT Model Explains Crowdfunding, Alternative Finance and Community Finance*. Available online: <https://ssrn.com/abstract=4302443> (accessed on 16 May 2024).
- Yu, L., W. K. Härdle, L. Borke, and T. Benshop. 2023. An AI Approach to Measuring Financial Risk. *The Singapore Economic Review* 68: 1529–49. <https://doi.org/10.1142/S0217590819500668>.

- Yue, T., D. Au, C. C. Au, and K. Y. Iu. 2023. Democratizing financial knowledge with ChatGPT by OpenAI: Unleashing the Power of Technology. Available online: <http://dx.doi.org/10.2139/ssrn.4346152> (accessed on 11 May 2024).
- Zaremba, A., and E. Demir. 2023. ChatGPT: Unlocking the Future of NLP in Finance. ChatGPT: Unlocking the future of NLP in finance. *Modern Finance* 1: 93–98. <https://doi.org/10.61351/mf.v1i1.43>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.