

Large Language Models and Generative AI in Finance: An Analysis of ChatGPT, Bard, and Bing AI

David S. Krause

Emeritus Professor, Finance Department, Marquette University, Milwaukee, WI, USA

david.krause@marquette.edu

Abstract

The field of finance is constantly evolving, and artificial intelligence (AI) is playing an increasingly important role in this evolution. Large language models (LLMs) and generative AI are two of the most promising AI technologies for the finance industry. This paper examines the application of LLMs and generative AI in the field of finance. It specifically focuses on three AI models: ChatGPT, Bard, and Bing AI. These models have shown significant potential in enhancing financial analysis, automating tasks, and providing valuable insights. By leveraging their language generation capabilities, they streamline financial analysis processes, improve decision-making, and advance natural language processing in finance. The analysis highlights the strengths and limitations of each model and emphasizes the importance of effective prompting and validation of AI-generated outputs. The evaluation of AI models in finance can be conducted using metrics such as ROUGE, BLEU, accuracy, precision, and perplexity. Despite their capabilities, challenges such as contextual understanding, factual accuracy, bias, fairness, and domain adaptation need to be addressed. Future directions for generative AI models in finance include domain adaptation, knowledge integration, explainability, interpretability, and ethical considerations. Generative AI models can be utilized responsibly and effectively in the finance industry; however, this will require careful consideration of the factors discussed in the paper.

Keywords: large language model, LLM, generative AI, ChatGPT, Bard, Bing AI, finance, financial analysis, investment analysis

Subject classification codes: C44, C45, C63, G14, G17

Large Language Models and Generative AI in Finance: An Analysis of ChatGPT, Bard, and Bing AI

Introduction

Financial analysis relies on accurate decision-making processes in various domains, including investment management, banking, risk assessment, and market forecasting. The recent advancements in artificial intelligence (AI) have sparked a growing interest in utilizing AI models to enhance financial analysis. Consequently, large language models (LLMs) and generative AI have emerged as powerful tools capable of generating text that closely resembles human language and improving various aspects of financial analysis.

Large language models are a type of AI model specifically designed to understand, generate, and predict human language. By leveraging statistical patterns and relationships learned from extensive text data, LLMs can generate coherent and contextually relevant sentences, making them invaluable for a wide range of natural language processing (NLP) tasks. Within the financial discipline, LLMs have been applied to machine translation, text generation, sentiment analysis, and more. Their ability to process and generate text has opened up new possibilities for enhancing financial analysis and decision-making.

Generative AI, a broader category of artificial intelligence, focuses on creating new data instances based on learned patterns and rules. In finance, generative AI models have the potential to revolutionize traditional financial analysis methods by quickly evaluating massive amounts of numerical and textual content, as well as providing unique insights. By synthesizing large volumes of financial data, generative AI models can assist analysts, traders, and investors in generating reports, predicting market trends, and optimizing portfolios. They offer a fresh perspective on financial analysis by complementing and expanding upon traditional research methods.

ChatGPT, Bard, and Bing AI

Three prominent examples of generative AI models that have demonstrated exceptional performance in language generation tasks are ChatGPT, Bard, and Bing AI. These AI models have undergone extensive training on vast text datasets, enabling them to provide contextually appropriate responses to various queries and conversations. Each has its own strengths and limitations, and the choice of the most helpful one depends on the specific task or inquiry.

Goode (2023) compares the three AI software tools and finds that they offer high level interactive chat experiences; however, they are prone to biases, often generating nonsensical or offensive responses. The review emphasizes the need for understanding the limitations of the models and the importance of considering biases when using them. Concerns are raised about truthfulness, bias, hallucinations, and the rapid evolution of AI technology, highlighting the need for a framework to regulate its use.¹

Dreibelbis (2023) conducts a test of the three AI models using the same set of questions and found that each produced a distinct response. ChatGPT relies on a curated dataset provided and lacks access to real-time internet information compared to the others. The comparative analysis underscores the evolving nature of AI chatbots and the need for greater transparency regarding the data underlying the AI systems.²

ChatGPT. Developed by OpenAI, ChatGPT is trained using the GPT (Generative Pre-trained Transformer) architecture and focuses on generating conversational

¹ Hallucinations in generative AI models involve the production of fabricated or factually unsupported outputs resembling human language, which can stem from biases, contextual limitations, or insufficient validation and fact-checking.

² The data for these models is unknown but is estimated to be in the petabyte range, which is equal to 1,000 terabytes or 1,000,000 gigabytes of data.

responses. It is trained on an extensive text corpus which enables it to understand context and provide coherent and contextually relevant answers, making it well-suited for chatbot applications in finance. OpenAI's ChatGPT made a decisive impact on the market with its launch in November 2022, quickly garnering more than 100 million users within two months, setting a record as the fastest technology application to achieve this milestone.³ In early 2023, OpenAI unveiled an upgraded version of ChatGPT, capitalizing on GPT-4, an improved large language model.⁴ This iteration generates text that is virtually indistinguishable from human-written content.

Bard. A recent innovation from Google, Bard was unveiled to the public in early 2023, following the amazing growth of ChatGPT's user base. Designed to emulate ChatGPT, Bard is a generative AI chatbot developed to cater to a wide range of needs including customer service, marketing, and education.

At its core, Bard is an expansive language model that has been thoroughly trained on an extensive corpus comprising text and code. Its vast array of capabilities includes generating text, facilitating language translation, crafting creative content, and providing informative responses to user queries. Anchoring Bard's functionality is LaMDA, an acronym for Language Model for Dialogue Applications, a powerful neural network with immense potential. LaMDA has undergone rigorous training using a diverse dataset encompassing books, articles, code, social media, and other forms of textual content. This comprehensive training equips Bard with the proficiency to deliver sophisticated and

³ This achievement shattered the previous record held by TikTok, which took nine months to reach the 100 million user count. However, Meta's app, Threads, reached 100 million users in just 5 days after its launch in July 2023.

⁴ GPT-4 has been trained on a massive dataset of text and code, including books, articles, websites, and social media posts. It is constantly being updated with new data.

adaptable language-based services. Bard is considered to be better at understanding and responding to factual queries, while GPT-4 is better at generating creative text.

Bing AI. In early 2023, Microsoft unveiled Bing AI, an AI-powered search engine that utilizes advanced techniques to provide relevant results. It understands query context and tailors outcomes based on users' search history. Powered by OpenAI's Prometheus Model 1, Bing AI combines the comprehensive Bing index, ranking, and answers with the creative reasoning capabilities of advanced GPT models. While Bing AI is considered to be accurate, relevant, and personalized, another strength is that it is integrated with other Microsoft products.

These developments serve as a testament to the growing interest in generative AI chatbots and indicate a promising trajectory of their adoption across various sectors in the years ahead. While initially designed for language-related tasks, the three models can be adapted and applied to diverse generative AI applications in finance, unlocking new possibilities for financial analysis and decision-making.

This paper aims to delve into the possibilities offered by generative AI models based on large language models in enhancing financial analysis. Our objective is to explore the capabilities of these models, assess their performance through an investment analysis case study, and analyze their influence on improving decision-making processes. Through an examination of the strengths and limitations of these models, our aim is to provide insights into their practical applications and make a valuable contribution to the burgeoning field of AI-powered financial analysis.

Literature Review of Generative AI in Finance

The application of generative AI models in the field of finance recently has garnered significant attention and has been the subject of several studies. Researchers have

explored the potential of generative AI models in various finance-related tasks, including financial document summarization, investment analysis, report generation, sentiment analysis, and portfolio optimization.

In the work by Osterrieder (2023), a comprehensive examination is conducted on the historical background and contemporary applications of AI in the financial services industry. Through the use of case examples, the author acknowledges the inherent challenges associated with integrating AI, including considerations of data quality, privacy, fairness, and ethics. The review underscores the significance of effective data management, appropriate model selection, and continuous evaluation as crucial factors for the successful implementation of AI in the realm of finance.

Hamadi et al. (2022) propose an approach to optimize portfolios by efficiently matching financial advisors to investors using a Generative Adversarial Network (GAN). They model the problem as an automated crowdsourcing platform and train the GAN to generate “ideal” financial advisor profiles based on investor inputs. Through extensive simulations, the AI-proposed solution demonstrates improved overall investment returns compared to baseline approaches. Future work includes validation on real-world datasets and enhancements to the GAN framework by incorporating additional features and an artificial intelligence model for risk minimization and portfolio optimization.

In the realm of financial document summarization, studies have demonstrated the efficacy of generative AI models in automatically generating concise and informative summaries from lengthy financial reports and articles. These models have shown promise in extracting key information, identifying critical insights, and presenting them in a coherent and understandable manner to analysts and investors.

Krause (2023a) offers a comprehensive guide to prompt usage in generative AI for financial analysis, encompassing the significance of prompts and offering effective

strategies. The article examines the effectiveness of various AI models in conducting a financial analysis of publicly traded firms.

Eckerli and Osterrieder (2021) provide an overview of using GANS in finance. They demonstrate that these can effectively generate time series and other financial performance data. While GANs present promising opportunities for data-driven modelling in quantitative finance, further research is needed to refine training, establish evaluation metrics, and improve stability.

Dalmasso et al. (2021) propose PayVAE, a generative model that learns the temporal and relational structure of financial transaction data. By applying PayVAE to a peer-to-peer payments dataset, they demonstrate its ability to generate realistic transactions. While the model captures the timing and occurrence patterns of simulated data, it has some limitations in generating diverse samples of active accounts. Nevertheless, the study presents a novel generative model that addresses the time-based and relational aspects of transaction data, making it a valuable contribution to banking and credit analysis.

Similarly, generative AI models have been employed in financial report generation. Researchers have leveraged the speed and power of these models to automate the production of accurate and comprehensive financial reports, eliminating the need for manual preparation. The generated reports encompass key financial metrics, analysis of financial statements, and insights into market trends, assisting analysts and decision-makers in making informed choices.

Krause (2023b) examines the risks associated with generative AI tools in the financial industry and proposes risk mitigation strategies. The article addresses concerns such as intellectual property infringement, offensive content generation, data security, fraud, and hallucinations. To mitigate these risks, financial firms should employ

generative AI tools on closed networks, use secure training data, implement robust security measures, provide employee training, and actively monitor output. Collaboration with regulatory bodies and industry associations, along with strict policies and partnerships with trusted vendors, can contribute to responsible usage.

Sentiment analysis, which is an important component of financial analysis, has been enhanced through generative AI models.⁵ These models have demonstrated the ability to analyze market sentiments and predict trends by processing vast amounts of textual data, including news articles, social media posts, management reports, and expert opinions. The utilization of generative AI models has improved the accuracy and efficiency of sentiment analysis in finance and other business applications.⁶

Li et al. (2023) addresses the gap in research regarding the performance of large language models on financial text analytics. They evaluate the capabilities of several AI models on various financial tasks, such as sentiment analysis and named entity recognition, using different financial datasets. The study reveals that while AI models excel in numerical reasoning tasks, they struggle with domain-specific knowledge and terminology, showing limitations in tasks requiring higher-level financial expertise.

Ömer and Karaarslan (2023) discuss the rise of AI models in business applications. They emphasize the need to consider the limitations of these tools, such as dataset weaknesses and algorithmic challenges, while acknowledging the evolving nature

⁵ Sentiment analysis is used to analyze textual data from sources like news articles and social media to understand market sentiment, make informed investment decisions, and identify potential opportunities or risks.

⁶ Sentiment analysis, which has a deep research history, helps analysts make informed investment decisions, identify risks, and improve customer satisfaction.

of generative AI. The study provides insights into user expectations, competitor analysis, and the current landscape of AI chatbots, drawing from a variety of sources.

Ahmed et al. (2023) conducted an extensive survey that compares ChatGPT and Bard. Their study delves into the architectures, training methodologies, performance evaluations, limitations, and ethical considerations of these models. The researchers envision a future where AI models play a central role in the daily activities of most participants in the financial industry.

In summary, the research on the application of generative AI models in finance is still in its early stages. Studies are beginning to shed light on the various architectures, training methodologies, performance evaluations, limitations, and ethical considerations of models like ChatGPT, Bard, and Bing AI. However, further research is needed to fully understand and explore the potential benefits and risks associated with the use of generative AI models in financial analysis.

Case Study

The following case study provides insights into the practical usability of ChatGPT, Bard, and Bing AI for investment analysis. While generative AI models offer various advantages, this case study notes the limitations and areas for improvement. It also highlights the importance of proper prompting.⁷

Effective prompting of generative AI in financial analysis is vital for achieving accurate and insightful results. By mastering prompt usage, financial analysts can enhance their decision-making and make informed investment choices. Effective prompts, tailored to AI chatbots like ChatGPT, Google Bard, and Bing AI, improve the

⁷ Prompting is the act of providing specific instructions or cues to an AI model to guide its behavior and generate desired outputs.

quality and accuracy of AI-generated responses. Best practices include treating chatbots as intelligent entities, incorporating context awareness, personalizing prompts, and using appropriate formats and phrasing.

The following example is a poorly written and ineffective prompt: “Give me information on Apple.” This prompt lacks specificity and does not provide clear guidance on what exactly is required in terms of financial analysis.⁸ It is too broad and open-ended, making it difficult to provide a focused and comprehensive response. There are other instances of prompt phrasing that can be confusing or ambiguous. For example, using the word “not” (a negation) often can confuse an AI system about what is being asked and will typically result in poor outcomes.

A stronger and more effective prompt is: “Compare the historical performance and financial ratios of Apple Inc. and Microsoft Corporation over the past five years, highlighting key differences in profitability, liquidity, and investment potential.” This prompt is specific and clearly outlines the task at hand, which is to conduct a comparative analysis of two specific companies. It specifies the period (past five years) and highlights the specific financial metrics (profitability, liquidity, and investment potential) to be analyzed. This prompts the AI model to provide a detailed and structured financial analysis, allowing for a more meaningful and comprehensive response.

Appendix A provides a comprehensive overview of the responses generated by the three AI models in relation to the question. Each model possesses its own set of strengths and weaknesses in terms of their respective responses. Notably, ChatGPT excels in time series ratio analysis, offering valuable insights in this aspect. However, it should

⁸ Ensuring precise prompts when seeking information on “Apple” is crucial to obtain relevant results about Apple Inc. rather than the fruit itself.

be noted that the data utilized by ChatGPT is relatively outdated, which can be a problem for financial analysts.⁹ On the other hand, Bard AI presents current information and demonstrates strong analytical capabilities, although it falls short in terms of providing a complete analysis spanning the full five-year period. Lastly, Bing AI displays robust financial comparative analysis and engages in dialogue effectively. However, it is worth mentioning that obtaining the desired information from Bing AI often requires the input of multiple prompts. By evaluating these distinct qualities, financial professionals can make informed decisions on which AI model best aligns with their specific needs.

Evaluation of the AI Models. The responses from ChatGPT, Bard, and Bing AI offer a comprehensive analysis of the historical performance and financial ratios of Apple Inc. and Microsoft Corporation. All of the models cover essential metrics related to profitability, liquidity, and investment potential, providing a well-rounded perspective on the companies' financial performance. The AI models demonstrate a level of robustness by delivering relevant financial information and analysis in response to the given query. They highlight the ability to generate accurate insights and present key metrics that contribute to a comprehensive understanding of the companies' financial positions.

It is important to note that the ChatGPT model's response is based on historical data up until the knowledge cutoff date in September 2021. In contrast, Bard and Bing AI offer a more up-to-date analysis, taking into account recent financial performance and market trends. Therefore, for the most timely analysis, relying on the responses from Bard AI and Bing AI would be more appropriate. Overall, the responses from the models present accurate historical financial data and ratios for the firms in the case study.

⁹ ChatGPT was trained on data up to September 2021, so it may not be able to answer questions about current events or provide information about things that have happened since then.

In summary, while each of the AI models offer a comprehensive analysis, it is important to acknowledge the limitations in terms of timeliness and accuracy due to the reliance on historical data. To ensure the most accurate and current information, it is advisable to complement the AI-generated insights with recent financial updates and market analysis from reliable sources. This approach will provide a more complete understanding of the companies' latest financial performance and investment potential.

Enhancing AI Model Performance

To enhance financial analysis using AI models, several strategies can be employed. Firstly, utilizing specific and detailed prompts can yield more accurate and relevant results. Instead of generic prompts, specifying the desired information and metrics of interest can provide more targeted analysis. For example, a prompt such as “Compare the gross and net profit margins of Apple Inc. and Microsoft Corporation for fiscal years 2018 to 2022, and identify significant trends and variations in profitability,” directs the AI models to focus on the specific analysis required.

Incorporating external data sources is another approach to improve analysis. AI models can benefit from accessing reliable external sources such as financial databases or reputable research platforms. Cross-referencing and validating information provided by AI models with trusted sources enhances the accuracy and reliability of the analysis. Reputable free online sources like Yahoo Finance, Google Finance, SEC EDGAR database, financial news websites (e.g., CNBC, Bloomberg, Reuters, Financial Times), and official company websites offer valuable financial data for this purpose.

Implementing fact-checking mechanisms can help verify the accuracy of AI-generated information. By comparing the results with established financial reports,

filings, or authoritative sources, any discrepancies or inconsistencies can be identified, helping to ensure the reliability of the analysis.

Continuous training of AI models with up-to-date financial data enables adaptation to changing market conditions and the incorporation of the latest financial performance.¹⁰ This enhances the timeliness and accuracy of the generated insights. Additionally, incorporating human expertise is crucial. Investment professionals can review AI-generated outputs, critically evaluate the information, and validate its relevance and accuracy. Their domain knowledge and experience provide valuable insights and ensure the AI-generated analysis aligns with industry standards.

Overall, improvements in financial analysis with AI models involves refining prompts, incorporating external data sources, implementing fact-checking mechanisms, continuous training, and leveraging human expertise. By combining the capabilities of AI models with human oversight and rigorous validation processes, the analysis becomes more robust, accurate, and reliable, leading to better-informed investment decisions.¹¹

Evaluation of AI-Generated Analysis

When assessing the performance of AI models in finance-related tasks, it is essential to employ relevant metrics to evaluate the quality and effectiveness of their generated outputs. Several commonly used metrics serve as useful quantitative measures to gauge AI model performance against specific objectives.

¹⁰ Models should be updated quarterly following the release of financial statements to ensure accuracy and reflect changes in the company's performance and operating environment.

¹¹ Human analysts will likely continue to play a vital role despite increased AI usage by providing oversight, validation, and risk mitigation to address the limitations and biases of AI models.

One widely used metric is ROUGE (Recall-Oriented Understudy for Gisting Evaluation), which measures the overlap between system-generated output and a human-generated reference summary.¹² By comparing the content, ROUGE captures the recall aspect of summarization tasks, providing insights into the quality of the AI-generated output. Appendix B contains an example of an evaluation of the case study AI-generated output using the ROUGE metric.

Another important metric is BLEU (Bilingual Evaluation Understudy), which evaluates the quality of machine-generated translations. It also compares a model's output with human-generated translations, assessing the n-gram overlap between them.¹³ BLEU offers a measure of the generated text's fluency and similarity to human-generated translations.

For sentiment analysis tasks, metrics like accuracy and precision are commonly used. Accuracy measures the percentage of correctly classified sentiments, while precision represents the proportion of true positive sentiment predictions out of all predicted positive sentiments. These metrics provide a quantitative assessment of model performance in sentiment analysis.¹⁴

In language modeling, perplexity is a commonly employed metric to evaluate a model's ability to predict the next word in a sequence. Lower perplexity values indicate

¹² ROUGE was created to address the need for automated evaluation measures for text summarization by Chin-Yew Lin at the Microsoft Research Asia in Beijing, China.

¹³ In natural language processing, an n-gram is a contiguous sequence of n items from a sample of text, and when evaluating machine-generated translations, the n-gram overlap is assessed to measure the similarity between the model's output and human-generated translations.

¹⁴ Various libraries and tools exist that provide functions to calculate these metrics based on the predicted sentiment labels and the "ground truth" labels.

better performance, indicating the model's proficiency in assigning higher probabilities to the correct words.

By leveraging these metrics, researchers and practitioners can effectively evaluate the performance of AI models in finance-related tasks. These metrics provide quantitative insights into a model's summarization capabilities, translation quality, sentiment analysis accuracy, and language modeling proficiency, enabling a comprehensive assessment of their performance, and guiding further improvements in the field.

Limitations and Challenges with AI Models

While ChatGPT, Bard, and Bing AI demonstrate promising capabilities in finance-related tasks, there are several limitations and challenges that should be considered. Firstly, these AI models can sometimes struggle with maintaining consistent context over longer conversations or documents, leading to incoherent responses or sensitivity to slight changes in input phrasing. Ensuring accurate contextual understanding and maintaining consistency throughout interactions pose ongoing challenges.¹⁵

Next, although generative AI models can produce fluent and contextually relevant text, there is a risk of producing outputs that are factually incorrect or misleading. The model's training on large-scale text corpora may inadvertently generate plausible but inaccurate information, necessitating careful validation and fact-checking processes.

Generative AI models are prone to biases present in the training data. If the data used for training contains biased or skewed information, it can manifest in the generated

¹⁵ Krause (2023a) asserts that the utilization of prompt templates can enhance the performance of a natural language generation model by directing it towards the specific task, enhancing the coherence and fluency of the generated text.

outputs. Achieving fairness and mitigating biases in the finance domain, where unbiased and objective analysis is crucial, remains a significant challenge.

While pre-training and fine-tuning can improve a model's performance on specific tasks, adapting it to different finance subdomains is still a challenge. Finance encompasses various specialized terminologies, regulations, and nuanced contexts, requiring adequate domain-specific training data and meticulous fine-tuning strategies.

In the evaluation of AI models in finance, metrics such as ROUGE, BLEU, accuracy, precision, and perplexity can be utilized. Comparative analyses across different tasks help identify variations in performance due to architectural differences between the models. Nevertheless, it is essential to acknowledge the challenges associated with maintaining context, ensuring factual accuracy, addressing biases, and achieving domain adaptation when applying AI models in finance.

Discussion and Future Direction

ChatGPT, Bard, and Bing AI are at the early stage of demonstrating their contributions to the field of finance. However, as identified by the researchers, leveraging language generation abilities will continue to enhance financial analysis, automate tasks, and provide valuable insights. The existing AI models are making improvements in automation in financial analysis processes, streamlining tasks such as summarization, report generation, sentiment analysis, and portfolio optimization. They will continue to help enable professionals and investors access critical financial information efficiently, resulting in more informed decision-making.

The advancements in natural language processing brought about by AI-based models have pushed the boundaries of language generation across various sectors. These models have shown their capability to understand and generate human-like text,

contributing to advancements in education, healthcare, cybersecurity, finance, insurance, and other disciplines.

There are several areas for further exploration and improvement within finance. Enhanced domain adaptation is one such area where research can focus on tailoring generative AI models to specific finance subdomains, ensuring accurate and reliable outputs within industry-specific terminologies, regulations, and nuanced contexts. Integrating domain-specific knowledge into these models can also enhance their understanding and generate more insightful outputs, incorporating financial expertise, economic principles, and market trends.

Explainability and interpretability are crucial considerations for the future development of generative AI models in finance. Building models that can provide transparency and insights into their decision-making processes will help establish trust and facilitate the integration of these models into critical financial systems. Additionally, ethical considerations must be addressed, including bias and fairness in outputs, data privacy and security, accountability, and the need for human-machine collaboration. Striking a balance between automated processes and human expertise is necessary to ensure responsible decision-making and accountability.

Addressing these future directions and ethical considerations requires interdisciplinary collaboration among researchers, practitioners, policymakers, and regulators. Establishing guidelines, frameworks, and standards for the responsible use of generative AI models in finance will ensure their ethical implementation and maximize their potential benefits while mitigating risks.

Conclusion

This paper has explored the application of large language models and generative AI, specifically ChatGPT, Bard, and Bing AI, in the field of finance. These AI models have

demonstrated their potential to enhance financial analysis, automate tasks, and provide valuable insights. By leveraging their language generation capabilities, they have streamlined financial analysis processes, improved decision-making, and advanced the field of natural language processing in finance.

The analysis of ChatGPT, Bard, and Bing AI has highlighted their individual strengths and limitations. Each model offers unique characteristics and performance attributes. ChatGPT excels in conversational abilities, Bard provides a wide range of language-based services with the LaMDA neural network, and Bing AI combines the Bing search engine with advanced GPT models for relevant financial analysis.

The case study conducted in this paper demonstrated the practical usability of these AI models for investment analysis. It emphasized the importance of effective prompting and the need to validate AI-generated outputs with external sources for accuracy and timeliness. Improving AI model performance in financial analysis requires refining prompts, incorporating external data sources, implementing fact-checking mechanisms, continuous training, and leveraging human expertise.

The evaluation of AI models in finance can be conducted using metrics such as ROUGE, BLEU, accuracy, precision, and perplexity. These metrics provide quantitative insights into the quality and effectiveness of the AI-generated outputs, guiding further improvements and assessment.

Despite the capabilities of these AI models, it is important to consider their limitations and challenges when applying them in finance. Contextual understanding, factual accuracy, bias and fairness, and domain adaptation present ongoing challenges. Addressing these challenges requires interdisciplinary collaboration and the establishment of guidelines and standards for the responsible use of generative AI models in finance.

Looking ahead, future directions for the development of generative AI models in finance include enhanced domain adaptation, knowledge integration, explainability and interpretability, and addressing ethical considerations. These areas of focus, along with interdisciplinary collaboration, will maximize the potential of generative AI models in finance while ensuring their responsible and ethical implementation.

Generative AI models, such as ChatGPT, Bard, and Bing AI, offer significant potential for enhancing financial analysis and decision-making in the finance industry. These models improve efficiency, automation, and insights in various finance-related tasks. However, their application requires careful consideration of limitations, adherence to ethical considerations, and ongoing efforts to address challenges. By harnessing the capabilities of these AI models while remaining mindful of their limitations, the finance industry can leverage their power to make more informed and effective financial decisions.

References

- Ahmed, Imtiaz, Mashrafi Kajol, Uzma Hasan, and Partha Protim Datta. "ChatGPT vs. Bard: A Comparative Study." TechRxiv, <https://www.techrxiv.org>. (2023).
- Aydın, Ömer, and Enis Karaarslan. "Is ChatGPT Leading Generative AI? What is Beyond Expectations?" *Available at SSRN 4341500* (2023).
- Dalmaso, Niccolo, Robert E. Tillman, Prashant Reddy, and Manuela Veloso. "PayVAE: A generative model for financial transactions." In AAAI 2021 Workshop on Knowledge Discovery from Unstructured Data in Financial Services Workshop. (2021).
- Dreibelbis, Emily. "ChatGPT vs. Google Bard vs. Bing: Which AI Chatbot Gives the Best Answers?" PCMag. <https://www.pcmag.com/news/chatgpt-vs-google-bard-vs-microsoft-bing-which-ai-chatbot-gives-best-answers>. (March 23, 2023).
- Eckerli, Florian, and Joerg Osterrieder. "Generative Adversarial Networks in Finance: An Overview." arXiv preprint arXiv:2106.06364 (2021).
- Goode, Lauren. "Review: We Put ChatGPT, Bing Chat, and Bard to the Test." Wired. <https://www.wired.com/story/review-ai-chatbots-bing-bard-chat-gpt/>. (March 30, 2023)
- Hamadi, Raby, Hakim Ghazzai, and Yehia Massoud. "A Generative Adversarial Network for Financial Advisor Recruitment in Smart Crowdsourcing Platforms." *Applied Sciences* 12, no. 19 (2022): 9830.
- Krause, David. "Proper Generative AI Prompting for Financial Analysis." *Available at SSRN 4453664* (2023).
- Krause, David. "Mitigating Risks for Financial Firms Using Generative AI Tools." *Available at SSRN 4452600* (2023).
- Li, Xianzhi, Xiaodan Zhu, Zhiqiang Ma, Xiaomo Liu, and Sameena Shah. "Are ChatGPT and GPT-4 General-Purpose Solvers for Financial Text Analytics? An Examination on Several Typical Tasks." arXiv preprint arXiv:2305.05862 (2023).
- Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." Paper presented at the Workshop on Text Summarization Branches Out, Barcelona, Spain. pp. 74-81. (2004).
- Osterrieder, Joerg. "A Primer on Artificial Intelligence and Machine Learning for the Financial Services Industry." *Available at SSRN 4349078*. (2023).

Appendix A

ChatGPT response to “Compare the historical performance and financial ratios of Apple Inc. and Microsoft Corporation over the past five years, highlighting key differences in profitability, liquidity, and investment potential.”

To provide a comprehensive analysis of the historical performance and financial ratios of Apple Inc. and Microsoft Corporation over the past five years, I will present a comparison of their profitability, liquidity, and investment potential. Please note that the following information is based on historical data up until my knowledge cutoff in September 2021.

Profitability:

1. Net Profit Margin (NPM):

- Apple:
 - 2020: 21.69%
 - 2019: 21.24%
 - 2018: 22.41%
 - 2017: 21.21%
 - 2016: 21.20%
- Microsoft:
 - 2020: 32.54%
 - 2019: 31.47%
 - 2018: 23.44%
 - 2017: 25.51%
 - 2016: 20.68%

Microsoft has generally displayed higher net profit margins compared to Apple, indicating better profitability over the past five years.

2. Return on Assets (ROA):

- Apple:
 - 2020: 12.67%
 - 2019: 14.01%
 - 2018: 16.71%
 - 2017: 14.78%
 - 2016: 14.13%
- Microsoft:
 - 2020: 15.57%
 - 2019: 13.80%
 - 2018: 11.90%
 - 2017: 11.49%
 - 2016: 9.88%

Apple has generally shown higher return on assets compared to Microsoft, indicating better utilization of its assets to generate profits.

Liquidity:

1. Current Ratio:

- Apple:
 - 2020: 1.36
 - 2019: 1.50
 - 2018: 1.12
 - 2017: 1.33
 - 2016: 1.35
- Microsoft:
 - 2020: 2.74
 - 2019: 2.67
 - 2018: 2.79
 - 2017: 2.60
 - 2016: 2.85

Microsoft consistently maintains a higher current ratio compared to Apple, indicating better short-term liquidity and ability to meet obligations.

2. Quick Ratio:

- Apple:
 - 2020: 1.01
 - 2019: 1.25
 - 2018: 0.93
 - 2017: 1.15
 - 2016: 1.24
- Microsoft:
 - 2020: 2.69
 - 2019: 2.62
 - 2018: 2.76
 - 2017: 2.57
 - 2016: 2.81

Similar to the current ratio, Microsoft maintains a higher quick ratio than Apple, indicating better short-term liquidity and ability to meet obligations without relying on inventory.

Investment Potential:

1. Price-to-Earnings (P/E) Ratio:

- Apple:
 - 2020: 34.99
 - 2019: 20.16
 - 2018: 17.86
 - 2017: 18.33

- 2016: 12.80
- Microsoft:
 - 2020: 35.76
 - 2019: 30.68
 - 2018: 25.06
 - 2017: 28.89
 - 2016: 26.85

Both Apple and Microsoft have relatively high P/E ratios, indicating that investors have higher expectations for future earnings growth.

2. Dividend Yield:

- Apple:
 - 2020: 0.71%
 - 2019: 1.21%
 - 2018: 1.44%
 - 2017: 1.57%
 - 2016: 1.98%
- Microsoft:
 - 2020: 0.90%
 - 2019: 1.44%
 - 2018: 1.68%
 - 2017: 1.97%
 - 2016: 2.34%

Both companies offer dividends, but Apple has generally had a lower dividend yield compared to Microsoft.

Overall, Microsoft has displayed better profitability and liquidity ratios over the past five years, while Apple has shown higher return on assets. Both companies have investment potential, as indicated by their relatively high P/E ratios, but Apple has had a lower dividend yield. It is important to conduct further research and analysis, including considering recent financial updates and market conditions, to make informed investment decisions.

Bard AI response to “Compare the historical performance and financial ratios of Apple Inc. and Microsoft Corporation over the past five years, highlighting key differences in profitability, liquidity, and investment potential.”

Here is a comparison of the historical performance and financial ratios of Apple Inc. and Microsoft Corporation over the past five years, highlighting key differences in profitability, liquidity, and investment potential:

Metric	Apple Inc.	Microsoft Corporation
Revenue	\$365.82 billion (2022)	\$168.09 billion (2022)
Net income	\$94.68 billion (2022)	\$68.95 billion (2022)
Profit margin	25.8% (2022)	41.1% (2022)
Return on equity	65.6% (2022)	43.6% (2022)
Current ratio	1.91 (2022)	1.32 (2022)
Quick ratio	1.12 (2022)	0.85 (2022)
Debt-to-equity ratio	0.38 (2022)	0.27 (2022)
Price-to-earnings ratio	23.29 (2023-02-14)	25.90 (2023-02-14)
Price-to-book ratio	6.42 (2023-02-14)	5.78 (2023-02-14)

As you can see, Apple Inc. has been more profitable than Microsoft Corporation over the past five years. Apple Inc. also has a higher profit margin, return on equity, and price-to-earnings ratio. However, Microsoft Corporation has a higher current ratio, quick ratio, and debt-to-equity ratio.

Overall, Apple Inc. appears to be a more profitable company than Microsoft Corporation. However, Microsoft Corporation may be a better investment for investors who are looking for a more liquid company with a lower debt load. Here are some additional details about each company:

- Apple Inc. is a multinational technology company that designs, develops, and sells consumer electronics, computer software, and online services. Apple is one of the Big Five American information technology companies, with Amazon, Google, Meta, and Microsoft being the others.
- Microsoft Corporation is an American multinational technology company that develops, manufactures, licenses, supports, and sells computer software, consumer electronics, personal computers, and related services. Microsoft is one of the Big Five American information technology companies, with Amazon, Apple, Google, and Meta being the others.

Both Apple Inc. and Microsoft Corporation are large, successful companies with a long history of profitability. However, there are some key differences between the two companies. Apple Inc. is more focused on consumer electronics, while Microsoft Corporation is more focused on software and services. Apple Inc. also has a higher profit margin and return on equity, while Microsoft Corporation has a higher current ratio, quick ratio, and debt-to-equity ratio. Ultimately, the best investment for you will depend on your individual investment goals and risk tolerance. If you are looking for a company with a long history of profitability and a high profit margin, Apple Inc. may be a good option for you. If you are looking for a company with a strong balance sheet and a low debt load, Microsoft Corporation may be a better option for you.

Bing AI response to “Compare the historical performance and financial ratios of Apple Inc. and Microsoft Corporation over the past five years, highlighting key differences in profitability, liquidity, and investment potential.”

Over the past five years, Microsoft has increased its gross profit on an annual basis by \$57.48 billion (98.47%) at an annual growth rate of 14.78%. In contrast, Apple has increased its annual revenue by \$131.52 billion (57.37%) at an average annual growth rate of 10.42%.

Source: Conversation with Bing AI, 7/12/2023

- (1) Microsoft (MSFT) Vs. Apple (AAPL): Updated Comparison After H1 2021.
<https://seekingalpha.com/article/4452489-microsoft-apple-updated-comparison-first-half-2021>.
- (2) Microsoft Financial Ratios for Analysis 2009-2023 | MSFT.
<https://www.macrotrends.net/stocks/charts/MSFT/microsoft/financial-ratios>.
- (3) Apple Financial Ratios for Analysis 2009-2023 | AAPL.
<https://www.macrotrends.net/stocks/charts/AAPL/apple/financial-ratios>.

In the past five years, both Apple and Microsoft have reported record-breaking profits. Apple made a \$21.7bn profit for the three-month period that ended in June, its best fiscal third quarter in its 45-year history, boosted by strong sales of the iPhone 12 and growth in its services business. Microsoft has also consistently generated high EBITDA margins of 47.99% combined with a 5Y average ROIC of 25%. In FY22, Microsoft generated \$198B in revenue, representing a growth of 18% compared to the previous year with the firm’s total revenues having grown a whopping 106% since FY17.

Source: Conversation with Bing, 7/12/2023

- (1) Google, Apple, and Microsoft report record-breaking profits.
<https://www.theguardian.com/technology/2021/jul/27/google-apple-and-microsoft-to-report-record-breaking-profits>.
- (2) Microsoft: Profitability Powerhouse Set for Long-Term Gains.
<https://seekingalpha.com/article/4583237-microsoft-profitability-powerhouse-set-for-long-term-gains>.
- (3) Microsoft (MSFT) Vs. Apple (AAPL): Updated Comparison After H1 2021.
<https://seekingalpha.com/article/4452489-microsoft-apple-updated-comparison-first-half-2021>.

(4) Better Buy: Apple vs. Microsoft | The Motley Fool.

<https://www.fool.com/investing/2021/03/13/better-buy-apple-vs-microsoft/>.

(5) Apple Inc. (NASDAQ: AAPL) | Analysis of Profitability Ratios. <https://www.stock-analysis-on.net/NASDAQ/Company/Apple-Inc/Ratios/Profitability>.

Apple holds \$195.6 billion in cash on hand, offset by \$112 billion in various forms of debt. Over the trailing 12 months, it generated \$80.2 billion in free cash flow. In comparison, Microsoft holds about \$132 billion in cash, against \$60.5 billion in total debt. Free cash flow for the company came in at about \$50.4 billion over the previous 12 months.

Source: Conversation with Bing AI, 7/12/2023

(1) Better Buy: Apple vs. Microsoft | The Motley Fool.

<https://www.fool.com/investing/2021/03/13/better-buy-apple-vs-microsoft/>.

(2) Apple Inc. (NASDAQ: AAPL) | Analysis of Liquidity Ratios.

<https://www.stock-analysis-on.net/NASDAQ/Company/Apple-Inc/Ratios/Liquidity>.

(3) Apple Current Ratio 2010-2023 | AAPL | MacroTrends.

<https://www.macrotrends.net/stocks/charts/AAPL/apple/current-ratio.net/NASDAQ/Company/Apple-Inc/Ratios/Profitability>.

Both Apple and Microsoft have been experiencing revenue growth despite inflation and other macroeconomic headwinds. Apple's iconic hardware products remain popular, hitting record adoption levels in its fiscal third quarter. Microsoft's cloud offerings helped the company accelerate revenue growth over the last two fiscal years. In terms of stock performance, earlier investors in MSFT stock have had every reason to celebrate over the past five years. Shares have risen by over 245%, from \$72.68 in July 2017 to \$253.21 at the time of writing on 2 December 2022.

Source: Conversation with Bing AI, 7/12/2023

(1) Apple vs Microsoft: Which Stock Is the Better Long-Term Investment

<https://www.fool.com/investing/2022/09/07/apple-vs-microsoft-which-stock-is-the-better-long/>.

(2) Microsoft Stock Price in 5 Years | Microsoft 5 Year Stock Forecast.

<https://capital.com/microsoft-stock-five-year-forecast>.

- (3) Will Microsoft Be Worth More Than Apple by 2025?.

<https://www.fool.com/investing/2022/09/23/will-microsoft-be-worth-more-than-apple-by-2025/>.

- (4) Today in Apple history: Microsoft investment helps save Apple - Cult of Mac. <https://www.cultofmac.com/567497/microsoft-investment-saves-apple/>.

Apple has announced a massive plan to become carbon-neutral across its entire business, including manufacturing, by 2030. It also announced a new recycling robot that will extract rare-earth metals from one of the most fragile systems in the iPhone. Apple has also been integrating augmented reality into its devices for years and it looks like the company will jump right into the territory of Meta, Microsoft, and Magic Leap with a long-expected mixed-reality headset in 2023.

Microsoft has been expanding its cloud-based services while reducing its dependence on desktop software. Microsoft developed more mobile apps for iOS and Android, launched new Surface devices, expanded its Xbox gaming business, and abandoned its struggling Windows Phone platform.

Source: Conversation with Bing AI, 5/18/2023

- (1) Apple Sets Climate Goals for 2030, Joins Amazon and Microsoft.

<https://www.wired.com/story/apple-sets-climate-goals-for-2030/>.

- (2) Apple's Mixed Reality Headset: What to Expect in 2023 - CNET.

<https://www.cnet.com/tech/computing/apple-ar-vr-headset-rumors-2023/>.

- (3) Tim Cook Reveals Apple's 10-Year Plan for Future Tech. <https://futurism.com/tim-cook-reveals-apples-10-year-plan-for-future-tech>.

- (4) Apple, Google and Microsoft Commit to Expanded Support for FIDO

<https://fidoalliance.org/apple-google-and-microsoft-commit-to-expanded-support-for-fido-standard-to-accelerate-availability-of-passwordless-sign-ins/>.

- (5) The Metaverse Set Off A Battle Between Tech Giants Google, Apple

<https://www.forbes.com/sites/jackkelly/2022/01/21/the-metaverse-set-off-a-battle-between-tech-giants-google-apple-microsoft-and-meta-to-build-virtual-and-augmented-reality-headsets/>.

Appendix B

Example of ROUGE Metric for Evaluating AI Model Performance

When evaluating the performance of AI models in summarizing the historical performance and financial ratios of Apple Inc. and Microsoft Corporation, the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metric was used.

ROUGE measures the overlap between system-generated summaries and human-generated reference summaries, providing insights into the accuracy and effectiveness of the generated summaries. By comparing the AI-generated content, ROUGE assesses the quality of the summaries and captures the recall aspect of summarization tasks. For example, this metric enables analysts to evaluate the performance of AI models in accurately summarizing the key information related to the historical performance and financial ratios of the companies, ensuring the quality and reliability of the generated outputs.

Performing a ROUGE analysis involves several steps. Firstly, a set of reference summaries, representing the desired quality and accuracy, is collected from known sources. Secondly, an output summary is generated using an AI model. Both reference and output summaries are then preprocessed to focus on meaningful content by removing stop words and punctuation. ROUGE scores are then computed by comparing the system-generated summaries with the reference summaries. The results are interpreted by analyzing the ROUGE scores, where higher scores indicate a greater overlap and similarity between the system and reference summaries, suggesting better performance in capturing the desired content. Finally, the ROUGE scores obtained from different AI models can be compared to assess their relative performance and identify areas for improvement.

Example ROUGE Analysis Results

As an example, the application of the ROUGE metric using Bard for summarizing the historical performance and financial ratios of Apple Inc. and Microsoft Corporation (see Appendix A) resulted in a ROUGE score of 0.82. This score indicates an 82% overlap between the AI system-generated summary and the reference summary, suggesting a relatively high level of similarity in the choice of words and phrases.

It is important to note that this example provides a basic illustration of how the ROUGE score can be used to quantitatively assess the quality and alignment of an AI system-generated summary versus a reference summary. Further analysis and comparisons with other AI models can provide deeper insights into their performance in summarizing financial data.

