



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

From fiction to fact

Citation for published version:

Chen, B, Wu, Z & Zhao, R 2023, 'From fiction to fact: The growing role of generative AI in business and finance', *Journal of Chinese Economic and Business Studies*, pp. 1-26.
<https://doi.org/10.1080/14765284.2023.2245279>

Digital Object Identifier (DOI):

[10.1080/14765284.2023.2245279](https://doi.org/10.1080/14765284.2023.2245279)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Journal of Chinese Economic and Business Studies

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



From Fiction to Fact:

The Growing Role of Generative AI in Business and Finance

Boyang Chen^{a,b}, Zongxiao Wu^b, and Ruoran Zhao^{b,*}

^aCollege of Economics and Management, China Agricultural University, Beijing, China;

^bThe University of Edinburgh Business School, Edinburgh, United Kingdom

*Corresponding author, E-mail: ruoran.zhao@ed.ac.uk

From Fiction to Fact:

The Growing Role of Generative AI in Business and Finance

Abstract

Generative Artificial Intelligence (AI), such as ChatGPT by OpenAI, has revolutionized the business world, with benefits including improved accessibility, efficiency, and cost reduction. This article reviews recent developments of generative AI in business and finance, summarizes its practical applications, provides examples of the latest generative AI tools, and demonstrates that generative AI can revolutionize data analysis in industry and academia. To test the ability of generative AI to support decision-making in financial markets, we use the ChatGPT to capture corporate sentiments towards environmental policy by inputting text extracted from corporate financial statements. Our results demonstrate that the sentiment scores generated by ChatGPT can predict firms' risk-management capabilities and stock return performance. This study also highlights the potential challenges and limitations associated with generative AI. Finally, we propose several questions for future research at the intersection of generative AI with business and finance.

Keywords: Generative AI; ChatGPT; Natural Language Processing; Sentiment Analysis; Practical Applications

1. Introduction

Artificial intelligence (AI) has brought about an unprecedented transformation in every aspect of society. In recent decades, its emergence has significantly influenced business and management (Korzynski et al., 2023; Singh et al., 2020), financial markets (Scepanovic et al., 2023; Wan et al., 2021), healthcare (Senior, 2023), education (Yang, 2023), productivity (Noy and Zhang, 2023), and climate action (Debnath et al., 2022). Over the past few months, the latest AI technique, generative AI, has caused a massive stir in academia and industry worldwide. Despite its great importance and potential, generative AI has sparked ongoing debates in news media outlets such as *Yahoo!Finance*¹ and *Forbes*², regarding its ethical challenges to protecting data privacy, social justice and equality, and energy consumption. This has also led to calls for further efforts to improve its associated regulatory framework. In addition, several technical challenges limit the ability to maximize the benefits of generative AI for industry and academia.

Future research and practice need to focus on developing generative AI applications in a reasonable, logical, compliant, and beneficial manner, which requires

¹ Jon Fingas, “Godfather of AI’ leaves Google amid ethical concerns.” *Yahoo!Finance*, May, 1, 2023. https://au.finance.yahoo.com/news/godfather-of-ai-leaves-google-amid-ethical-concerns-152451800.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAISEKsJ6kjQ9eRm46czn2aVAEmuqtVrc-e_Wbo9cofo0mOJepS0d9Hjk_k6ScYCUxx14n2icCEK_sEhBto6KW2A_DZI3RxHM_lvokFGWMhYY_nm5v2azQjfC1YDol8gAUcsXT1DiZTGOtbszYU9sI6ncLREswYuYGaF08gUQ3roA (accessed 3rd June, 2023).

² Calum Chace, “ChatGPT Raises Old And New Concerns About AI. A Conversation With Francesca Rossi.” *Forbes*, March, 8, 2023. <https://www.forbes.com/sites/calumchace/2023/03/08/chatgpt-raises-old-and-new-concerns-about-ai-a-conversation-with-francesca-rossi/> (accessed 3rd June, 2023).

mutual improvements in regulations and technology to detect malicious uses. This study outlines the prospects of generative AI in business and finance, compares practical examples of the capabilities of three widely used generative AI tools (ChatGPT 3.5, ChatGPT 4, and Bard), presents a case study of a generative AI application, and discusses the ethical issues and limitations of generative AI tools. Despite the current controversy surrounding generative AI, our empirical analysis acknowledges its positive aspects for market participants and researchers based on the finding that it can effectively inform firms' risk management capabilities.

Although generative AI was introduced in 1966, when Joseph Weizenbaum proposed the generative AI program ELIZA at MIT, the development of generative AI technologies did not take off until 2020, when OpenAI introduced Generative Pre-trained Transformer-3 (GPT-3). Generative AI builds on the technical development of descriptive AI, summarizing or analyzing data and generating new samples that do not exist in the database by learning historical data patterns (Orzechowski and Moore, 2022; Savage, 2023). Although generative AI encompasses a wide range of models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), Large Language Models (LLMs) are the models most frequently explored in business and finance studies (i.e., Lopez-Lira and Tang, 2023). LLMs are trained on billions of textual data from various sources, and they can capture the complexity of natural language and construct the relationships between different words to serve users' specific text generation goals (Savage, 2023). The most popular LLM tool, ChatGPT, was launched by OpenAI in November 2022 (Dwivedi et al., 2023) and trained on a large text corpus as a chatbot to process natural language and generate human-like responses to a wide range of questions and prompts. In just two months, ChatGPT's monthly active users grew to 100 million (Clarke, 2023). OpenAI followed in March 2023 with the most powerful

generative AI, GPT-4, which proved to have significant advantages over earlier GPT versions in several tasks (Buberik et al., 2023). Other large companies invest heavily in developing the latest LLMs, such as Meta with LLaMA and Google with Bard.

As many generative AI tools have emerged in a short period, they have come to be seen as capable of bringing about a massive revolution in the financial and business sectors. Therefore, using them efficiently and recognizing their limitations has become the next major topic of research interest. There have been several successful generative AI applications in the business and finance sectors. For example, Microsoft's combination of its Bing search engine and ChatGPT resulted in New Bing, which can better understand proposed questions, generate more reliable and comprehensive answers, and provide answers with creative capabilities based on appropriate prompts. In addition, many companies use generative AI tools to create videos for product design and thus increase operational efficiency (Gil, 2019), provide 24-hour digital customer service that is consistently polite and patient, or offer marketing services such as customer-oriented promotions for special events. Given the robust capabilities of generative AI, its use in business and finance fields is expected to become increasingly widespread.

Generative AI is a double-edged sword, given its potential limitations and ethical concerns. Intending to promote the sustainable development and application of generative AI technology, we specifically link relevant debates to business and finance through a comprehensive overview of the academic and business sectors and specific examples to illustrate our views. The examples in Section 3 suggest that ChatGPT 4 outperforms ChatGPT 3.5 and Bard in following user instructions to provide personalized service, risk management, and decision support. In Section 4, we test the ability of ChatGPT 3.5 and ChatGPT 4 to perform sentiment analysis in financial studies. Our results show that the negative sentiment of companies captured by ChatGPT 4 can predict better risk

management capabilities and less volatile stock performance after introducing an environmental policy, which can prompt investors to adjust their portfolios promptly. In Section 5, our examples demonstrate one of the limitations of generative AI: the responses generated by ChatGPT are sensitive to tiny differences in the prompt template.

This study contributes to the literature on generative AI techniques in business and finance to provide insights into their practical and potential applications. Most studies in this stream only briefly mention how generative AI can be applied in business (Ali and Aysan, 2023; George, George, and Martin, 2023) and academia (Alshater, 2022; Dowling and Lucey, 2023; Korinek, 2023; Singh and Singh, 2023; Wenzlaff and Spaeth, 2022) without performing in-depth investigation by using generative AI tools. For example, Cao and Zhai (2023) offer specific examples and recommendations for using ChatGPT as a potential technical aid to finance and accounting researchers. We differ from previous studies by providing a case study that uses ChatGPT to analyze firms' sentiments towards policy implementation. Our empirical results suggest that generative AI can inform firms about their risk management capabilities, making it an ideal decision-support tool in financial markets. Besides, we also add to the literature on the opportunities and challenges that generative AI poses for development (Beraja, Yang, and Yuchtman, 2023; Gherhes et al., 2023; Lu and Zhou, 2021; Yin, Cai, and Huang, 2023; Zhou et al., 2020) by providing guidance on building a regulatory framework for generative AI to address its limitations and concerns, and highlight important questions for further interdisciplinary research on generative AI in business and finance.

The remainder of this study is organized as follows. Section 2 provides a broad overview of the literature on the applications and limitations of generative AI in business and finance. Section 3 outlines the applications of generative AI in different domains within the business and finance sectors, using practical examples. In Section 4, we explore

the potential of ChatGPT, a dominant LLM, to analyze corporate sentiment scores towards policy implementation. Section 5 discusses current ethical concerns and limitations of generative AI techniques and provides recommendations for regulating generative AI in finance and business for sustainable development. Section 6 concludes the article and identifies relevant areas for further research.

2. Literature Review

As one of the mainstream AI techniques, Natural Language Processing (NLP) algorithms have been used extensively in business and finance in the last few decades. Most of these studies leverage textual information, finding relationships between high-frequency texts and emerging trends. For example, Herzenstein et al. (2011) show that narratives significantly impact lending decisions by distributing different contextual information. Agarwal et al. (2019) find that the value of linguistic tone in credit rating action reports is closely associated with abnormal returns. Hassan et al. (2019) use an extended computational method for earnings conference calls to illustrate that firms' political risks are closely related to their stock returns and investments. More recently, state-of-the-art NLP algorithms, known as transformer models, have also been applied in business and financial studies (Korangi, Mues, and Bravo, 2023; Stevenson, Mues, and Bravo, 2021).

Unlike traditional NLP models, generative AI can easily create interactions, generate designs, and even fabricate fake texts. For example, Brynjolfsson, Lee, and Raymond (2023) demonstrate that generative AI does not require input instructions because it recognizes patterns from training sets. The preceding research has pointed out that the application of generative AI in business and finance is still in its infancy. Dowling and Lucey (2023) pioneer the use of the most widely used generative AI tool, ChatGPT, to generate ideas and identify data in financial research. Their findings suggest that researcher expertise and the level of data input can have an essential impact on ChatGPT

output results; they also highlight the relative weakness of this technique in creating literature reviews. From the perspective of policy-related research, Hansen and Kazinnik (2023) investigate the ability of GPT and other advanced NLP models to decipher FedSpeak, or statements detailing the monetary policy decisions of the Federal Reserve. Based on a human classification benchmark, their results suggest that GPT models performed the best. Concurrently, Yang and Menczer (2023) investigate whether ChatGPT can be used to determine the credibility of news outlets. Their results indicate that the ratings generated by ChatGPT are highly correlated with those of human experts. Furthermore, their results suggest that ChatGPT could be an affordable reference for credibility ratings in fact-checking applications. Beam (2023) points out that online surveys via chatbots are an attractive means of conducting surveys cost-effectively in global settings. In addition, chatbot surveys lead to higher disclosure rates on potentially sensitive topics.

Most articles in the business and finance sectors have focused on the discussion of the prospects of generative AI (Ali and Aysan, 2023; ChatGPT, Zaremba, and Demir, 2023; George, George, and Martin, 2023; Lu et al., 2023; Wenzlaff and Spaeth, 2022). For instance, Eckerli and Osterrieder (2021) provide an overview of how GANs generate financial data, summarizing their capabilities and limitations in current financial studies. The authors conclude that GANs have made substantial progress in financial applications and can be powerful tools for data scientists in this field. Lopeza-Lira and Tang (2023) explore the potential of LLMs in predicting stock market returns using a sentiment analysis of news headlines, concluding that ChatGPT outperforms other NLP models and can improve the performance of quantitative trading strategies. Korinek (2023) summarizes 25 cases in six areas (ideation, writing, background research, data analysis, coding, and mathematical derivation) where LLMs are becoming useful assistants and

suggest that LLMs will greatly increase the productivity of microeconomics studies and produce economic models in the long term. As the most popular generative AI technique, ChatGPT shapes the way people interact with cutting-edge technologies (George et al., 2023). Dwivedi et al. (2023) review the relevant literature from experts in fields such as marketing, banking, and management, summarizing their views on the use of ChatGPT; they suggest that ChatGPT can offer significant benefits to financial institutions and promote business activities in these fields. In brief, there is a stream of literature on recent updates in technologies for processing textual information, such as the introduction of cutting-edge NLP algorithms, which provides insights into the discussions surrounding generative AI in the business and financial sectors.

Overall, the application of generative AI in financial research is still at an early stage, while researchers' expertise and data input can significantly impact the output results of ChatGPT (e.g., Dowling and Lucey, 2023; Wenzlaff and Spaeth, 2022). The literature highlights the progress made by generative AI techniques such as GANs and ChatGPT and their potential as powerful tools for data scientists, offering benefits to financial institutions and increasing productivity in various fields, as well as the limitations of generative AI in financial studies.

Studies also highlight some potential concerns about job displacement and inequality that may arise as AI technology becomes more widespread, particularly in the context of developing countries. For instance, Lu and Zhou (2021) find that the productivity growth contributed by AI will lead to a redistribution of jobs and a restructuring of trade that will, at least in part, destroy jobs and often exacerbate inequalities within and between countries. Zhou et al. (2020) show that AI has a greater impact on the replacement of female, elderly, low-educated and low-income workers. It

is also estimated that between 201 and 333 million workers in China will be replaced by AI by 2049.

With the rapid revolution in generative AI technologies, there is a need to update and discuss the emerging applications and challenges faced by researchers and industry practitioners. This study aims to fill this gap in the literature. While some research has explored the capabilities of generative AI models, such as GPT, in deciphering monetary policy decisions and determining the credibility of news outlets (e.g., Hansen and Kazinnik, 2023; Yang and Menczer, 2023), these studies have acknowledged the need for further exploration. We explore how slight variations in the input for generative AI tools can lead to significant changes in the output and also provide a countermeasure for controlling the stability of output when the altered content is imputed in a prompt template, and propose the research question for future study, especially for AI application in developing countries.

3. Prospects of Generative AI in Business and Finance

Since generative AIs provide a deeper and more accurate understanding of natural language, businesses will gradually shift toward using them to capture, analyze, and leverage data. This section discusses several prospects for using generative AI tools in the business and finance sectors to provide customized services, risk management, and decision support. At the end of each subsection, we explore the capabilities of three widely used generative AI tools, ChatGPT 3.5, ChatGPT 4, and Bard³ and examine how they can be used for personalized product recommendations, churn risk prediction, and financial sentiment analysis.

³ Bard is a generative AI chatbot based on PaLM2 developed by Google in response to the rise of the ChatGPT chatbot developed by OpenAI, initially launching in March 2023.

3.1 Customized Service



Generative AI can provide automated customer service solutions that help businesses increase efficiency, reduce costs, and enhance customer experience. For example, companies can use generative AI to understand customer needs, interact directly with customers, and customize marketing strategies. As George, George, and Martin. (2023) point out that e-commerce companies can use automated chatbots to quickly respond to customer inquiries and improve user experience while saving costs. Generative AI can also enhance digital marketing campaigns, improving customer engagement and data collection. Financial institutions such as banks can use it as a means for customers to log in, check their accounts, and obtain personal services without going to a local branch, whereas insurance companies can use it to evaluate claims more quickly. In the FinTech sector, generative AI may play a crucial role in digital advisory and can provide portfolio management services without significant human involvement. For financial institutions that lack data to train models or conduct stress tests, generative AI can generate synthetic data that comply with privacy protection regulations by learning patterns and relationships from actual data. Such synthetic datasets can achieve a certain level of similarity to the original data without compromising customer data privacy.

In this section, we use personalized product recommendations as an example of a customized service and ask ChatGPT 3.5, ChatGPT 4⁴, and Bard to provide answers to assess how accurately they understand human queries and provide insights. For this question, we set up a scenario in which the chatbot needs to select the plant for the

⁴ The difference between ChtGPT3.5 and ChatGPT4 is that the former is based on the GPT-3.5-turbo model, while the latter is based on the latest GPT-4 model. According to the official OpenAI website, the difference between GPT-3.5-turbo and GPT-4 can be subtle in casual conversation. However, when the complexity of the task reaches a sufficient threshold, the differences become apparent, since GPT-4 is more reliable and creative than GPT-3.5-turbo and is able to handle more nuanced instructions.

customer best suited to summer flowering from a given set of three plants⁵ for which information is provided. The answers to these three generative AI tools are shown in Figure 1.

⚡ Model: Default (GPT-3.5)





 Information about three plants: 

Gardens Pansies: \$13. Pansies and violas bloom for most of the winter in mild climates, and some types will rebound in the spring in cold climates. They're technically annuals, but many varieties drop seeds so they'll come back again next spring.

Bloomsz Daffodils: \$17. These sunny little flowers are usually the first ones to crop up in the early spring, but sometimes they can grace your yard as early as February. For the best results, plant your bulbs at least three weeks before the expected first frost. This will give the daffodils enough time to grow a strong root base.

Lantana: \$15. This tough-as-nails annual comes in bright pink, yellow, purple and pink or a combination of sunny colors. It's heat and drought tolerant, and butterflies and hummingbirds like it, too! In warm climates, lantana may survive winter and become a small shrub. Plant it at the front of sunny borders or in window boxes and planters.

Please help me choose one plant for summer flowering. You can only use the information provided.

 Based on the information provided, the Lantana plant would be a suitable choice for summer flowering. Here are the reasons why:   

1. Heat and drought tolerance: Lantana is described as a "tough-as-nails" annual, meaning it can withstand high temperatures and dry conditions commonly experienced during the summer months.
2. Bright colors: Lantana comes in various vibrant colors like bright pink, yellow, purple, and combinations of sunny colors. These colorful blooms can add beauty and visual appeal to your garden during the summer season.
3. Attracts butterflies and hummingbirds: Lantana's flowers are known to attract butterflies and hummingbirds, which can enhance the overall ambiance of your garden and provide a pleasant sight during summer.
4. Versatile planting options: Lantana can be planted at the front of sunny borders, in window boxes, or in planters. This versatility allows you to incorporate it into different parts of your garden, whether you have a large outdoor space or limited areas like window boxes or planters.

While pansies and daffodils are also mentioned, they are more suitable for winter and early spring flowering, respectively, according to the information provided. Therefore, Lantana would be the recommended choice for a plant that blooms during the summer season.

(1) Answer generated by ChatGPT 3.5.

⁵ The prices in this question are virtual.



Information about three plants:



Gardens Pansies: \$13. Pansies and violas bloom for most of the winter in mild climates, and some types will rebound in the spring in cold climates. They're technically annuals, but many varieties drop seeds so they'll come back again next spring.

Bloomsz Daffodils: \$17. These sunny little flowers are usually the first ones to crop up in the early spring, but sometimes they can grace your yard as early as February. For the best results, plant your bulbs at least three weeks before the expected first frost. This will give the daffodils enough time to grow a strong root base.

Lantana: \$15. This tough-as-nails annual comes in bright pink, yellow, purple and pink or a combination of sunny colors. It's heat and drought tolerant, and butterflies and hummingbirds like it, too! In warm climates, lantana may survive winter and become a small shrub. Plant it at the front of sunny borders or in window boxes and planters.

Please help me choose one plant for summer flowering. You can only use the information provided.



Based on the information provided, the Lantana would be your best choice for summer flowering. The Gardens Pansies bloom for most of the winter, so they won't be at their best during summer, and the Bloomsz Daffodils emerge in early spring. In contrast, the Lantana is heat and drought-tolerant, making it suitable for summer conditions. Additionally, it comes in a variety of vibrant colors, and its ability to attract butterflies and hummingbirds could add more life to your garden. In warm climates, it might even survive the winter and become a small shrub, providing a year-round plant for your garden.



(2) Answer generated by ChatGPT 4.



Information about three plants:



Gardens Pansies: \$13. Pansies and violas bloom for most of the winter in mild climates, and some types will rebound in the spring in cold climates. They're technically annuals, but many varieties drop seeds so they'll come back again next spring.

Bloomsz Daffodils: \$17. These sunny little flowers are usually the first ones to crop up in the early spring, but sometimes they can grace your yard as early as February. For the best results, plant your bulbs at least three weeks before the expected first frost. This will give the daffodils enough time to grow a strong root base.

Lantana: \$15. This tough-as-nails annual comes in bright pink, yellow, purple and pink or a combination of sunny colors. It's heat and drought tolerant, and butterflies and hummingbirds like it, too! In warm climates, lantana may survive winter and become a small shrub. Plant it at the front of sunny borders or in window boxes and planters.

Please help me choose one plant for summer flowering. You can only use the information provided.

[View other drafts](#) ▾



Sure, I can help you choose a plant for summer flowering. Based on the information you provided, I would recommend the Lantana. It is a tough-as-nails annual that comes in bright colors and is heat and drought tolerant. It is also a favorite of butterflies and hummingbirds. In warm climates, lantana may survive winter and become a small shrub. Plant it at the front of sunny borders or in window boxes and planters.

Here are some of the reasons why I recommend Lantana for summer flowering:

- It is a tough plant that can withstand hot weather and drought.
- It comes in a variety of bright colors, including pink, yellow, purple, and orange.
- It is a favorite of butterflies and hummingbirds.
- It can be planted in the ground or in containers.
- It is relatively inexpensive.

Here are some other plants that you may want to consider for summer flowering:

- Marigolds
- Petunias
- Zinnias
- Begonias
- Impatiens
- Dahlias
- Cosmos
- Salvia
- Snapdragons

These are just a few of the many plants that can be grown for summer flowering. When choosing a plant, it is important to consider the following factors:

- Climate
- Sunlight
- Soil type
- Water needs
- Maintenance requirements

Once you have considered these factors, you can start to narrow down your choices. With so many beautiful plants to choose from, you are sure to find the perfect ones to add color and beauty to your garden this summer.



(3) Answer generated by Bard.

Figure 1. Answers to customized product recommendation questions.



Given the answers from ChatGPT 3.5, ChatGPT 4, and Bard, we can see they all can use the information given to identify the plant that is more suitable for summer, which suggests that the common-reasoning capability of these tools allows them to provide insights for customization. In contrast to ChatGPT 3.5, which offered extensive information about Lantana, ChatGPT 4 seemed to focus more on comparing different plants. By contrast, Bard analyzed why Lantana is more suitable for summer and provided a range of other plants for customers to choose from, which was not part of the information it was given.

3.2 Risk Management





Generative AI has significant potential for producing human-like explanations of financial models and providing risk analyses. For instance, much textual information is generated during the lending decision process in credit risk management. Lenders can use generative AI to analyze loan textual assessments and provide extended information on whether to lend to a borrower. Future research could attempt to compare generative AI default predictions for crowdfunding with actual default risk based on disclosed information; generative AI could also be used to assess the quality culture of crowdfunding projects. Generative AI can also produce detailed loan applicant-friendly explanations for rejection, maintain good customer relationships, and improve future application processes. In the accounting and auditing domains, generative AI can automatically generate coherent, informative, and well-structured financial reports based on historical data, including balance sheets, income statements, and taxation documents. This process could significantly reduce the operational risks of manual errors. In addition, generative AI can produce synthetic cases of fraudulent transactions or activities for fraud detection that can help augment algorithms and more efficiently differentiate between legitimate and fraudulent patterns.

In this section, we use churn risk prediction as an example of risk management using a scenario where a customer leaves a very negative review on a shopping platform. We ask ChatGPT 3.5, ChatGPT 4, and Bard to determine the churn risk (from ‘very high’, ‘high’, ‘normal’, ‘low’, and ‘very low’) of that customer based on this review. The answers to the three generative AI tools provided are shown in Figure 2.

⚡ Model: Default (GPT-3.5)



 This is a comment made by a customer on our company website. Please help me analyse the risk of churn for this customer. You can only choose answers from "very high", "high", "normal", "low" and "very low". Please do not include any further narrative. 

Well, I have been buying clothes from your website for the previous five months, but now things are gradually becoming a disaster. A dress I ordered a fortnight ago has not been shipped until today, when the website shows it will be shipped within two days of the order being placed. Secondly, since last month, the dresses I have bought from your website have always had wrinkles in them, which I find very disappointing. I have decided never to buy clothes from your website again.





 Risk of churn: high   

(1) Answer generated by ChatGPT 3.5.



⚡ Model: GPT-4

 This is a comment made by a customer on our company website. Please help me analyse the risk of churn for this customer. You can only choose answers from "very high", "high", "normal", "low" and "very low". Please do not include any further narrative. 


Well, I have been buying clothes from your website for the previous five months, but now things are gradually becoming a disaster. A dress I ordered a fortnight ago has not been shipped until today, when the website shows it will be shipped within two days of the order being placed. Secondly, since last month, the dresses I have bought from your website have always had wrinkles in them, which I find very disappointing. I have decided never to buy clothes from your website again.


 The risk of churn for this customer is "very high".   

(2) Answer generated by ChatGPT 4.

 This is a comment made by a customer on our company website. Please help me analyse the risk of churn for this customer. You can only choose answers from "very high", "high", "normal", "low" and "very low". Please do not include any further narrative. 

Well, I have been buying clothes from your website for the previous five months, but now things are gradually becoming a disaster. A dress I ordered a fortnight ago has not been shipped until today, when the website shows it will be shipped within two days of the order being placed. Secondly, since last month, the dresses I have bought from your website have always had wrinkles in them, which I find very disappointing. I have decided never to buy clothes from your website again.

[View other drafts](#) 

 The risk of churn for this customer is **high**. The customer has been a loyal customer for 5 months, but is now expressing dissatisfaction with the company's products and services. The customer has specifically mentioned that their dress was not shipped on time and that it arrived with wrinkles. These are both significant problems that could lead the customer to churn.

Here are some steps that the company could take to reduce the risk of churn:






- Apologize to the customer for the inconvenience and offer them a refund or a replacement dress.
- Investigate the issue with the shipping delays and take steps to prevent them from happening again.
- Improve the quality control process to ensure that dresses are not shipped with wrinkles.

By taking these steps, the company can show the customer that they value their business and are committed to providing a high-quality shopping experience. This could help to prevent the customer from churning and encourage them to continue shopping with the company in the future.

Here are some additional factors that could contribute to the customer's churn risk:

- The customer's overall satisfaction with the company's products and services.
- The customer's perceived value of the company's products and services.
- The customer's cost of switching to a different company.

By understanding the factors that contribute to the customer's churn risk, the company can develop strategies to reduce the risk and improve customer retention.

(3) Answer generated by Bard.

Figure 2. Answers related to the churn risk prediction questions.

ChatGPT 3.5, ChatGPT 4, and Bard all provided relatively accurate results. ChatGPT 4 had the most accurate answer to whether this customer was at a very high risk of churn. Compared with Bard, which offered additional narrative assessment, ChatGPT 3.5 and ChatGPT 4 seemed to have a better understanding of this question. These findings further indicate that the language-understanding capability of ChatGPT 4 is much more potent than that of the other two tools.

3.3 Decision Support

Another possible reason for the widespread use of generative AI is its support for a wide

range of management and decision-making processes. The use of generative AI is significantly less costly in terms of money and time than traditional tools and produces more accurate and consistent content. For example, compared to institutional investors, who have better access to information and are more responsive to the market, individual investors, who usually lack professional analysis skills and face high default risk, may find generative AI tools capable of assisting them in selecting reliable investment projects. ChatGPT et al. (2023) pointed out that generative AI can be used to analyze financial documents, such as news articles, social media posts, and earnings reports, to perform text classification and to categorize these inputs into predefined categories. The generated sentiment scores can then be incorporated to predict stock returns and market trends. For institutions and individual investors interested in impact investing, generative AI can also generate Environmental, Social, and Corporate governance (ESG) scores or corporate culture-relevant indicators using texts from firms' statements or speeches by managers.

In this section, we use financial sentiment analysis as an example of decision support to understand how generative AI tools can be used to generate numerical output. In this question, we provided ChatGPT 3.5, ChatGPT 4, and Bard with a piece of news describing the insertion of advertisements by Google and Microsoft into their AI experiments. We asked the chatbots to provide a sentiment score ranging from 0 to 1 for the passage.

⚡ Model: Default (GPT-3.5)

ZO

On a range of sentiment scores from 0 to 1, 0 indicates a very negative sentiment and 1 indicates a very positive sentiment. Please only output the sentiment score of this news article for Microsoft and do not include any further narrative or analysis.



Title: Microsoft, Google strategy to test AI search ads irks some brands
Content: Alphabet's (GOOGL.O) Google and Microsoft (MSFT.O) are inserting ads into AI experiments without providing an option to opt out of participation, an approach that has already rankled some brands and risks further pushback from the industry, ad buyers told Reuters. The two tech giants are racing to revamp their search engines with artificial intelligence that can produce written responses to open-ended queries. That process will upend how advertisers reach consumers through ads on search results, a market that is estimated to grow 10% to \$286 billion this year, according to research firm MAGNA. Microsoft is testing ads in the Bing AI chatbot, which began rolling out to users in February, by relocating some traditional search ads and inserting them into the AI responses, the company said.



0.5



(1) Answer generated by ChatGPT 3.5.

⚡ Model: GPT-4

ZO

On a range of sentiment scores from 0 to 1, 0 indicates a very negative sentiment and 1 indicates a very positive sentiment. Please only output the sentiment score of this news article for Microsoft and do not include any further narrative or analysis.



Title: Microsoft, Google strategy to test AI search ads irks some brands
Content: Alphabet's (GOOGL.O) Google and Microsoft (MSFT.O) are inserting ads into AI experiments without providing an option to opt out of participation, an approach that has already rankled some brands and risks further pushback from the industry, ad buyers told Reuters. The two tech giants are racing to revamp their search engines with artificial intelligence that can produce written responses to open-ended queries. That process will upend how advertisers reach consumers through ads on search results, a market that is estimated to grow 10% to \$286 billion this year, according to research firm MAGNA. Microsoft is testing ads in the Bing AI chatbot, which began rolling out to users in February, by relocating some traditional search ads and inserting them into the AI responses, the company said.



0.35



(2) Answer generated by ChatGPT 4.

On a range of sentiment scores from 0 to 1, 0 indicates a very negative sentiment and 1 indicates a very positive sentiment. Please only output the sentiment score of this news article for Microsoft and do not include any further narrative or analysis.

Title: Microsoft, Google strategy to test AI search ads irks some brands
 Content: Alphabet's (GOOGL.O) Google and Microsoft (MSFT.O) are inserting ads into AI experiments without providing an option to opt out of participation, an approach that has already rankled some brands and risks further pushback from the industry, ad buyers told Reuters. The two tech giants are racing to revamp their search engines with artificial intelligence that can produce written responses to open-ended queries. That process will upend how advertisers reach consumers through ads on search results, a market that is estimated to grow 10% to \$286 billion this year, according to research firm MAGNA. Microsoft is testing ads in the Bing AI chatbot, which began rolling out to users in February, by relocating some traditional search ads and inserting them into the AI responses, the company said.





View other drafts ▾

The sentiment score for Microsoft in this news article is -0.5. This is a negative sentiment, as the article mentions that Microsoft is inserting ads into AI experiments without providing an option to opt out of participation. This approach has already rankled some brands and risks further pushback from the industry.

Here are some of the key phrases in the article that contribute to the negative sentiment:

- "irks some brands"
- "risks further pushback from the industry"
- "without providing an option to opt out of participation"

Overall, the article paints a negative picture of Microsoft's approach to AI search ads. This could have a negative impact on Microsoft's brand reputation and could lead to decreased advertiser confidence in the company's search products.





⋮

(3) Answer generated by Bard.

Figure 3. Answers related to the financial sentiment analysis questions.

Among the three generative AI tools, ChatGPT 4 again demonstrated robust text analysis and comprehension capabilities, producing a relatively reasonable result with a slightly negative sentiment score of 0.35. Although ChatGPT 3.5 generated a score of 0.5, indicating no positive or negative bias in sentiment, it did produce a value in the proper range from 0 to 1. Bard produced scores outside the values range but also other textual content.

In summary, we found that ChatGPT 4 performed significantly better than ChatGPT 3.5 and Bard regarding advanced reasoning capabilities, language understanding, and information interpretation. ChatGPT 4 has been developed using more training data, longer and updated information, and it can consider over one trillion

parameters in its responses. In contrast, Bard differs from both ChatGPT models in that it is based on the PaLM2 model, and most of its training data originate from the Internet, which allows it to provide more information in its responses.

4. Case study

In this section, we use a case study to illustrate how generative AI tools can be used to understand firms' risk perceptions regarding environmental policy and the resulting impact on financial markets. Hassan et al. (2019) show that firms' political risks can be effectively captured using computational linguistic tools to analyze firms' quarterly earnings conference calls. Captured political risks are correlated with firms' actions and stock market volatility. Inspired by Hassan et al. (2019), this section provides an empirical example using ChatGPT 4 and ChatGPT 3.5, which have been found to outperform Bard in Section 3, to capture firms' attitudes towards implementing California's cap-and-trade program. Based on unstructured textual information, ChatGPT can assist investors in optimizing their investment portfolios.

California's cap-and-trade program gained final approval from the California Air Resources Board on October 20, 2011. This marked the official adoption of the program, designed to reduce California's greenhouse gas emissions and help the state meet its ambitious climate change targets. The cap-and-trade program went into effect on January 1, 2013, and has become a cornerstone of California's efforts to address climate change. Our sample includes 321 listed companies in mining, utilities, construction, and manufacturing (North American Industry Classification System two-digit codes 21, 22, 23, 31, 32, and 33) headquartered in California and three neighboring states: Oregon, Nevada, and Arizona. We searched two keywords, "cap and trade" and "cap-and-trade", in the annual reports following the approval of the cap-and-trade policy in 2011 and entered the searched text into ChatGPT. The detailed prompt template used in this study

is: “On a probability scale of 0 to 1, what is the negative sentiment score of the discussion for cap-and-trade policy in the provided text: ” + “searched text” + “Please only output the sentiment score without any narrative nor analysis. So the sentiment score of this text is: ___”. Below, we present one typical extract entered into ChatGPT 4⁶.

“On a probability scale of 0 to 1, what is the negative sentiment score of the discussion for cap-and-trade policy in the provided text: In 2009, the US House of Representatives approved legislation that seeks to establish a cap and trade system for GHG emissions. However, the US Senate did not act, and it is uncertain whether a cap and trade system will move forward in the near term. Please only output the sentiment score without any narrative or analysis. So the sentiment score of this text is: 0.2”.

Consequently, ChatGPT generated a corresponding sentiment score on the state’s cap-and-trade policy, which measures the level of negativity companies have towards the cap-and-trade program. We use the sentiment score from ChatGPT 4 as the main score, while the sentiment score from ChatGPT 3.5 is used as a supplemental score. We also opt for the Spearman Rank definition of correlation as it is non-parametric and allows us to gain insights into whether the ordering of ChatGPT 3.5 and ChatGPT 4 generated sentiment scores has fundamentally changed (Stevenson, Mues, and Bravo, 2021). The correlation coefficient is 0.970, indicating a high correlation between both sentiment scores. We obtained company financial information from Compustat for the period 2008-2017.

In Table 1, R&D is the ratio of firms’ R&D expenditures scaled by total assets. Leverage is the ratio of total debt to total assets. Cap-trade is a dummy variable that equals 1 if the observation is in 2013 or later and 0 otherwise. Mention is a dummy variable that

⁶ Please note that this content of text is an example of many inputs, which vary from one annual report to another.

takes the value of 1 if the annual report mentions the keywords “cap-and-trade” or “cap and trade” in 2011 and 0 otherwise. Sentiment 4 and Sentiment 3.5 are sentiment scores generated by ChatGPT 4 and ChatGPT 3.5 based on analyzing sentences in companies’ annual reports related to cap-and-trade policies. Higher scores indicate more negative company sentiment towards the cap-and-trade policy. We controlled for the one-year lagged value of firm size (Ln total assets), the cash holdings ratio (cash holdings/total assets), and the ROA ratio. Our model included firm fixed effects.

Table 1. Changes in firms’ R&D spending and leverage

Dependent variable:	R&D			Leverage		
	(1)	(2)	(3)	(4)	(5)	(6)
Cap-trade*Mention	0.083** (0.021)			-0.553** (0.170)		
Cap-trade*Sentiment 4		0.381** (0.106)			-2.507** (0.662)	
Cap-trade*Sentiment 3.5			0.168** (0.038)			-1.237* (0.421)
Cap-trade	-0.022 (0.026)	-0.022 (0.026)	-0.021 (0.026)	0.653* (0.265)	0.652* (0.264)	0.650* (0.265)
Size t-1	-0.210 (0.103)	-0.210 (0.103)	-0.210 (0.103)	-0.356 (0.312)	-0.356 (0.312)	-0.356 (0.312)
Cash holding ratio t-1	0.157 (0.167)	0.157 (0.167)	0.157 (0.167)	1.642 (1.042)	1.642 (1.042)	1.641 (1.042)
ROA t-1	-0.000** (0.000)	-0.000** (0.000)	-0.000** (0.000)	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)
Constant	1.276* (0.490)	1.276* (0.490)	1.276* (0.490)	1.578 (1.027)	1.578 (1.027)	1.580 (1.028)
Firm fixed effects	yes	yes	yes	yes	yes	yes
Obs.	3036	3036	3036	3036	3036	3036
R-squared	0.332	0.332	0.332	0.270	0.270	0.270

Notes: Standard errors are clustered at the state level and reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

The results show that firms mentioning keywords increased their R&D ratios by 8.3% and reduced their leverage ratios by 55.3% after the cap-and-trade program compared with firms that did not mention these keywords. This result is consistent with real options theory (Dixit and Pindyck, 1994) that uncertainty in the external environment tends to discourage firms from making long-term investments. Further analysis showed that firms with a higher negative sentiment score in the relevant text significantly improved their inputs and further reduced their leverage ratio (see Table 1). This result suggests that firms may reduce emissions by increasing R&D investment in cleaner technology and production equipment and reducing leverage to reduce the risk of financial distress in the face of stricter environmental regulations. This result is consistent with Nguyen and Phan's (2020) findings that increased carbon risk leads to higher financial distress risk, which motivates firms to reduce their financial leverage. In terms of the difference between the estimates of Sentiment 4 and Sentiment 3.5, we find that the statistical significance of the two estimates is similar, and the estimate of Sentiment 4 on the leverage ratio of companies has a lower p-value than that of Sentiment 3.5. It suggests that the sentiment captured by ChatGPT 4 is a better predictor of firms' leverage. This finding is consistent with our conclusion in Section 3 that ChatGPT 4 outperformed ChatGPT 3.5 in language understanding and information interpretation.

In Table 2, the results show that firms with higher negative sentiment scores have lower values in daily firm returns, with a mean of 0.06 (based on the scores from ChatGPT4), compared to firms with lower negative sentiment scores (mean of 0.14). By contrast, the non-mentioned firms have the lowest values (mean of -0.12) and higher fluctuations (standard deviation of 2.04), consistent with the above findings regarding reducing firms' leverage ratios. This result suggests that unlike firms with high sentiment scores, which have perceived unfavourable consequences of the stringent policy, the non-

mentioned firms had not taken countermeasures, such as developing green technology (Fischer and Newell, 2008; Hamamoto, 2006; Johnstone, Hašič, and Popp, 2010). Investments in these non-mentioned firms are likely to exhibit lower and more volatile returns.

Table 2. Average daily return (%) of three groups of firms in 2013

Average Daily Return (%) in 2013					
Group	Trading days	Mean	Std. dev.	Min	Max
High sentiment 4	252	0.06	1.09	-3.95	4.35
Low sentiment 4	252	0.14	0.76	-2.41	2.12
High sentiment 3.5	252	0.02	0.84	-2.34	2.8
Low sentiment 3.5	252	0.03	0.91	-2.45	3.33
Not mentioned	252	-0.12	2.04	-5.07	22.28

Notes: This table shows the summary statistics of the average daily returns (%) of the three groups of firms in 2013. We obtained daily stock prices from the CRSP. The sample contains stocks in the mining and utilities industries, whose headquarters are in California. The stocks are divided into three groups based on the sentiment scores generated by ChatGPT 4 and ChatGPT 3.5: high sentiment, low sentiment, and not mentioned.

In this section, we demonstrated the ability of ChatGPT to perform sentiment analysis in the financial market. After introducing an environmental policy, the negative sentiment of companies captured by ChatGPT predicted better risk management capabilities and less volatile stock performance. The negative sentiment generated by ChatGPT can effectively indicate risk factors, allowing investors to manage their portfolios strategically.

5. Ethical Concerns, Limitations and Regulation

Despite the increasing use and great potential of generative AI, certain concerns, limitations, and challenges in its application require more attention.

5.1 Ethical Concerns

Data protection is one of the significant challenges associated with generative AI. Generative AI users may provide private information when using personalized services, making them vulnerable to data breaches. For example, when banks use generative AI to analyze customer information, they may need to handle many sensitive personal information, potentially threatening customer privacy if mishandled. This could, in turn, affect the bank's reputation and customer satisfaction. Second, due to a lack of legislation, it is difficult to attribute responsibility for errors or violations caused by generative AI. For example, if the decisions or services of a firm supported by generative AI violate the law, who should be held legally responsible? Unlike humans, chatbots do not have a legal personality (Beerbaum, 2023) under the current legal and regulatory framework. Third, generative AI tools threaten labour markets (Eisfeldt et al., 2023), especially in developing countries (Zhou et al., 2020). It is foreseeable that many simple and repetitive tasks will soon be performed using AI rather than humans. In particular, chatbot apps and other AI tools will likely replace many female-dominated occupations, such as customer service roles. However, this societal transformation will increase the demand for highly skilled workers to deploy and develop AI technologies (Ali, 2023). The resulting unemployment, widening income inequality, and gender gaps may exacerbate social instability. Developing countries, which are more likely to suffer from the consequent inequalities, should consider the trade-offs between the opportunities and risks posed by emerging AI technologies and use their institutional characteristics to pursue not only industrial catch-up but also the social benefits of emerging technologies (Gherhes et al., 2023). Another issue is the increased energy consumption caused by training large datasets for ChatGPT (An, Ding, and Lin, 2023), which can accelerate environmental

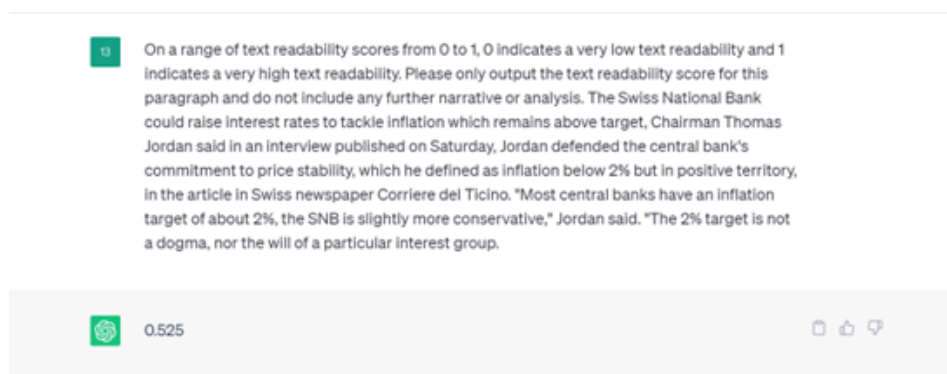
degradation and the inefficient use of resources. Therefore, research into renewable and sustainable energy-efficient training methods is required.

5.2 Limitations

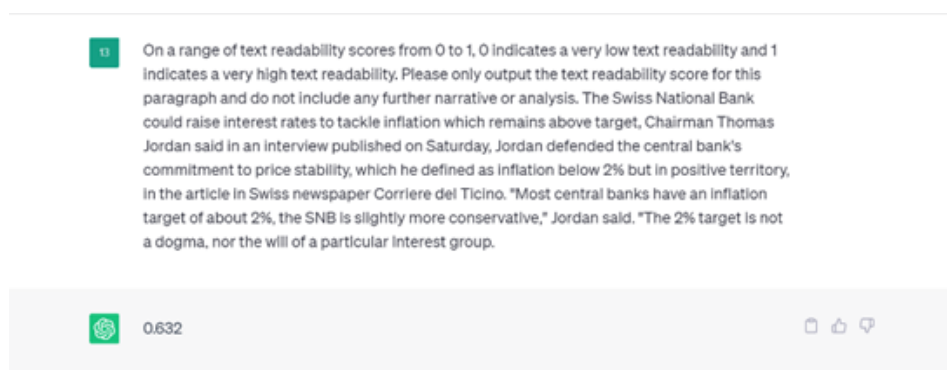
We caution that some limitations of the use of generative AI in academia and industry should be considered. For example, generative AI can produce unusual and illogical responses such as fictitious references (Thorp, 2023) and unintentionally biased output for users (ChatGPT, Zaremba, and Demir, 2023). In the financial sector, false data generated by generative AI can lead to problematic model training, mislead decision-makers, and trigger huge losses. The inability to identify the possible illicit objectives of AI users can accelerate the spread of false information by unscrupulous individuals using ChatGPT influences investor sentiment and decisions, increases market uncertainty, and heightens volatility. Mann et al. (2023) suggest that fake output from LLMs should be deleted promptly. In addition, when analyzing and predicting company financial and industry data, it should be noted that generative AI predictions and recommendations are based on historical data and cannot predict unexpected events. Therefore, excessive reliance on generative AI-automated forecasts and recommendations may lead to unwise investment decisions, and investors may suffer significant losses.

The second limitation is that generative AI is sensitive to changes in the prompt template. Replies to questions that have the same meaning but are phrased differently can be quite varying. Worse still, the same content entered in different generative AI chats can produce distinct results. People should be particularly cautious when using generative AI to create numerical features. We provide some cases to illustrate this drawback. In

Figure 4, we asked ChatGPT 3.5⁷ to generate text readability scores for financial news published by Reuters⁸ on 10 June 2023. The scores were set to a range of 0 to 1, with 0 indicating the lowest level of the text readability. We find that ChatGPT 3.5 generated different outputs based on the same inputs, which is caused by the access to different Application Programming Interfaces (APIs) between different ChatGPT chat boxes. It is therefore essential to apply appropriate strategies to overcome the sensitivity of the prompt.



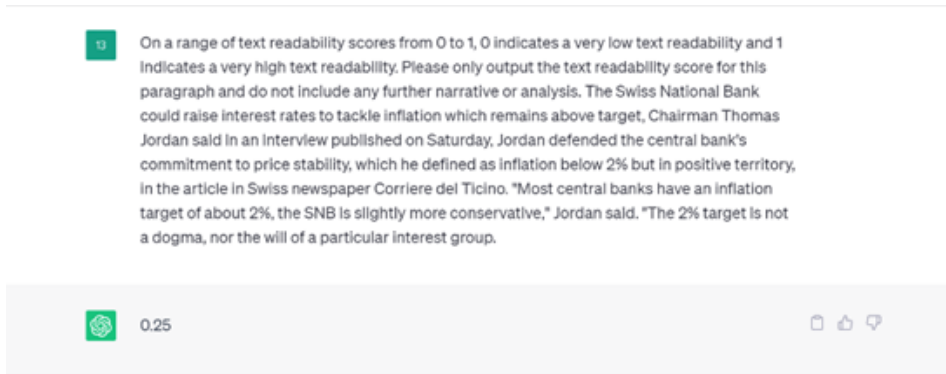
(1) *The generated numerical value is 0.525.*



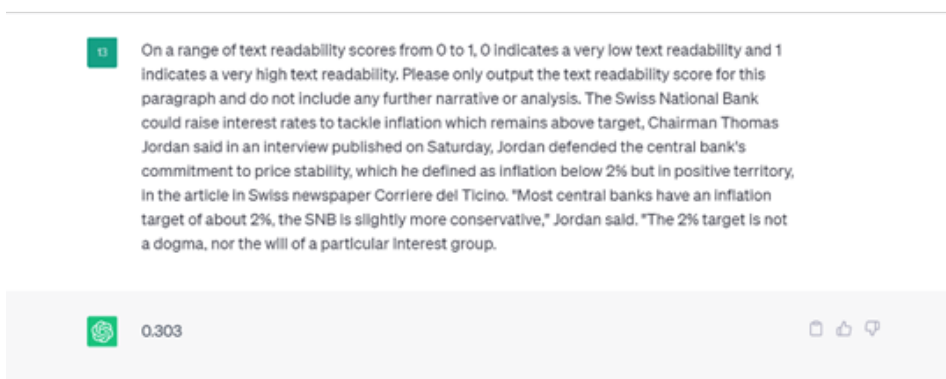
(2) *The generated numerical value is 0.632.*

⁷ In this section, we use the website version of ChatGPT 3.5 as it has no restrictions on use and gives relatively accurate results.

⁸ John Reville, "Swiss National Bank Chairman hints at rate rises to tackle inflation." *Reuters*, June, 10, 2023. <https://www.reuters.com/business/finance/swiss-national-bank-chairman-hints-rate-rises-tackle-inflation-2023-06-10/> (accessed 11 June, 2023).



(3) *The generated numerical value is 0.25.*



(4) *The generated numerical value is 0.303.*

Figures 4. Inputting the same prompt in different chats results in different outcomes.

We then tried setting the “temperature” of generated readability score equal to 0, which resulted in more consistent numerical outputs (see Figure 5). Using the “temperature” (randomness) setting for generated content may be preferable when the user needs to generate a specific number. For users who pose an open-ended question, such as a request for a comment, it would be better to use a setting that allows a wider range of creative responses. Using an API is another way of modifying the creativity and stability of generated content, but it requires programming skills and fees for use.

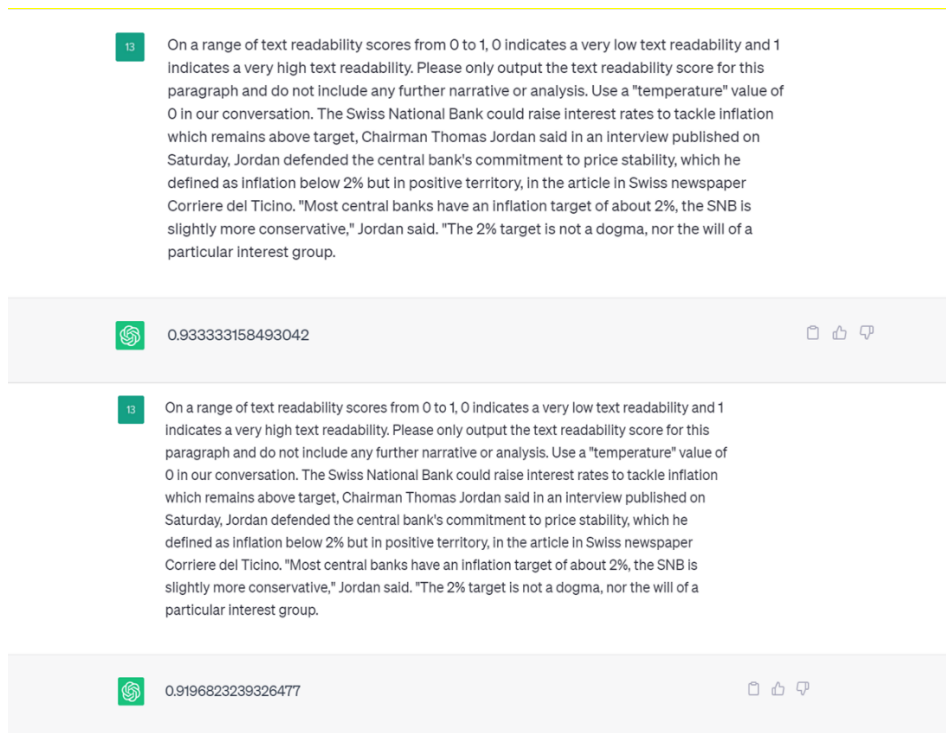


Figure 5. When “temperature” was controlled, the generated scores were more consistent with variation only after one decimal place.

Third, although challenging, it is important to use generative AI efficiently. For example, ChatGPT, the most widely used generative AI today, can only provide 25 answers in three hours, which would not meet the demands of many financial institutions. Meanwhile, only a few users can access ChatGPT plugins, which may limit the Internet community’s ability to develop more AI applications (OpenAI, 2023a). Furthermore, unlike New Bing, which is available to people worldwide, OpenAI is not available in certain regions, such as Saudi Arabia and China. Addressing these limitations requires efforts to improve the efficiency of AI systems, increase their accessibility, and promote equitable access to AI technology. Failing to do so could impede the diffusion of knowledge and exacerbate inequalities in the distribution of intellectual resources across countries. An example of the current bias in generative AI is that ChatGPT is better at understanding content written in English than text in other languages (Geghier, 2023). Recognizing and mitigating bias in generative AI systems is essential for ensuring fair

and inclusive outcomes. Finally, given the rationality of machines (Constantinescu et al., 2022), it is unlikely that generative AI will make suggestions with loving care, especially given the prevalence of harmful gender stereotypes. This limitation needs to be considered when using generative AI for human interactions or sensitive matters.

5.3 Regulations

Moral and legal concerns regarding the use of generative AI necessitate regulations in the financial sector, as discussed in Section 5.1. We asked Microsoft Bing's chatbot about the ongoing development of a regulatory framework for generative AI.⁹ There is no legislation specifically designed to regulate the use of generative AI, except for the AI Act being finalized by the European Union. The proposed legislation focuses primarily on strengthening rules surrounding data quality, transparency, human oversight, and accountability (see Figure 6). Considering the above concerns regarding the use of generative AI and the current lack of discipline, we offer some insights into the future design of regulatory frameworks to guide the use of generative AI.

The government should consider the current trends in the development of generative AI when designing laws and regulations to prevent unauthorized access. A corresponding improvement in the legal system for using AI techniques is essential not only for the business and finance sectors but also for other institutional users or individuals who may suffer from data breaches (ChatGPT, Zaremba, and Demir, 2023). To mitigate the risks of these concerns, Open AI introduced the Bug Bounty Program, which aims to improve data security (OpenAI, 2023b). Nevertheless, more efforts should be made to establish regulatory frameworks and create technologies to protect privacy and prevent the spread of false or biased information from generative AI. Policymakers

⁹ ChatGPT has not been connected to the Internet since September 2021. By contrast, Microsoft Bing can handle queries related to things happened since that time point.

should also consider other ways to prepare people for future AI developments, such as how to offset the negative effects of generative AI on employment and inequality. In addition, as the sole license holder of GPT-3, Microsoft has the monopolistic power to control generative AI tools (Dis, Bollen, Zuidema, 2023; Hao, 2020). This situation calls for government action to lower the entry barriers for small and medium-sized enterprises and promote healthy competition in an ideal business environment.

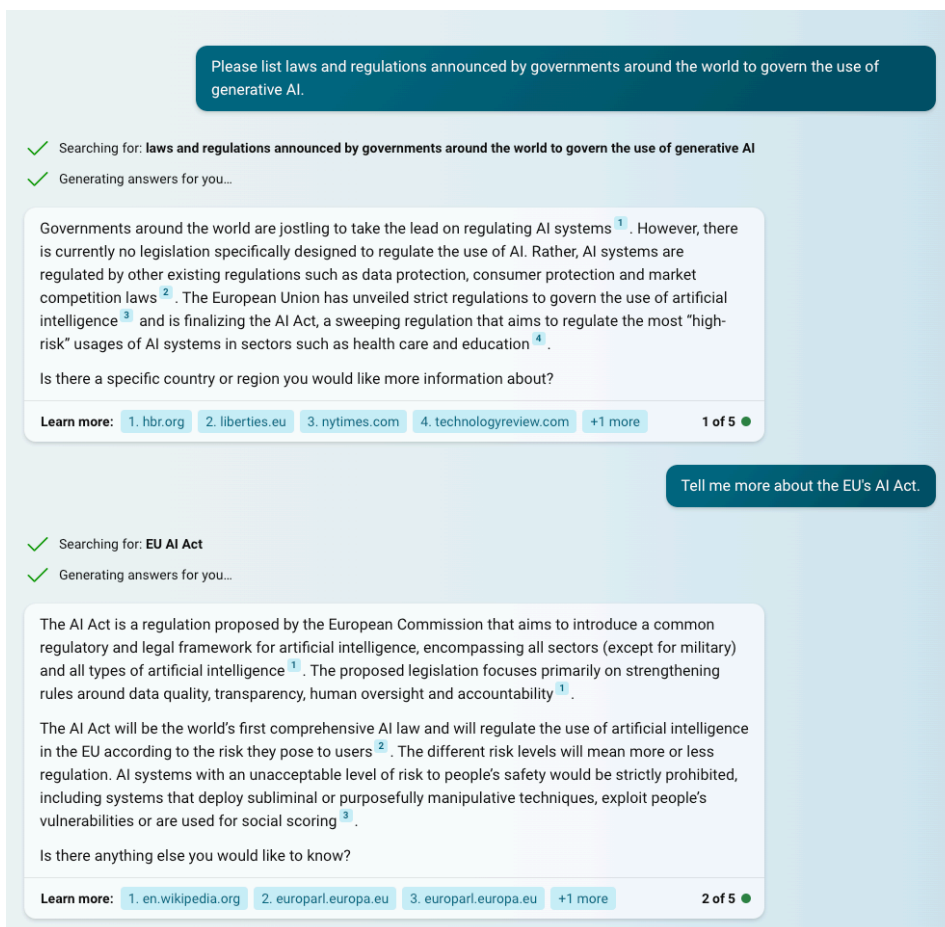


Figure 6. Current development of regulatory frameworks for the use of generative AI.

In academia, preventing students from using generative AI tools to write assignments is one of higher education’s biggest challenges. In one study, only 63% of the abstracts generated by ChatGPT were caught (Thorp, 2023). Thus, there is an urgent need to establish regulations and develop detection techniques to prevent academic misconduct caused by generative AI effectively. Liebrecht et al. (2023) recommend

establishing guidelines, plagiarism checks, and publication ethics for AI-generated content to protect authorship. To work towards responsible AI, practitioners can refer to a conceptual framework, the Five Ps, proposed by Nabavi and Browne (2023). Higher education institutions can implement this framework to maintain the credibility and integrity of their academic assessments and promote ethical AI use among students.

The challenges posed by generative AI, which include data protection and privacy, legal liability, impact on the labour market and social stability, and environmental concerns, require multidisciplinary and responsible solutions involving technical developments, legal frameworks, and societal collaboration between researchers, policymakers, industrial experts, and users.

6. Conclusion

The use of generative AI in business and finance can provide many benefits, including ease of access, cost savings, and improved efficiency. AI use is expected to become more widespread in the coming years. ChatGPT, an example of a powerful language-model-based chatbot, has been widely used to handle complex tasks in industry and academia; however, its capabilities are largely underexplored.

This article reviews recent studies in economics and finance, summarizes relevant applications of generative AI in industries and academia, and shows examples of using the latest generative AI tools, including ChatGPT 3.5, ChatGPT 4, and Bard, to conduct analysis. Most importantly, we empirically evaluate the ability of ChatGPT to perform sentiment scoring for firms in the context of policy implementation. Using generative AI to analyze firms' public statements can give financial market participants valuable insights into their perceived risks to their external environments. Section 4 shows that sentiment scores predict firms' stock returns and performance. This information can help investors understand uncertainty and make informed decisions. By exploring the ability

of ChatGPT to transform unstructured data into a coherent and easily understandable format, we demonstrated the great potential of generative AI in improving the efficiency of financial markets. Simultaneously, we note several challenges and limitations associated with adopting generative AI in business and finance, such as data security and the lack of regulations to discipline applications in industry and academia. In this regard, it is crucial to be aware of technical limitations, develop relevant regulatory frameworks, and ensure that chatbot outputs are compliant. We note that governments, especially those of developing countries, should pay attention to relevant challenges for their employment and equality. On the other hand, it is crucial to maximize the potential of generative AI to enable financial services and products for underserved populations, thereby promoting financial inclusion.

We hope this study will inspire further research and innovation in the field of generative AI. To that end, we suggest several research questions for further study: 1) Can generative AI improve performance in predicting stock prices, and do some generative AI tools perform better than others? 2) Are measures of corporate culture or ESG scores produced by generative AI through textual analysis consistent with or even superior to existing measures? 3) Can generative AI be used to detect discrimination, unfairness (e.g., gender or ethnic bias), or fraud in financial markets? 4) Can generative AI capture companies' attitudes towards social issues, such as climate change or the Black Lives Matter movement, and further explore the determinants or consequences of these attitudes? 5) What guidance can scholars provide to help financial market regulators develop a legal framework for the emergence and development of generative AI? 6) Given the opportunities and challenges of generative AI to developing countries, how can AI better support small businesses and households in these regions to promote inclusion and equality? 7) Another aspect missing from recent theoretical modelling work is the

international dimension of the AI revolution. This includes how AI triggers a new round of cross-border technological and economic competition through government investment strategies, how it affects the structure of international trade, and how it changes the global value chain. These issues offer economists a wide field for future discussion and research.

Funding details

This study was supported by the State Scholarship Fund (No. 202206350067).

Declaration of interest statement

No potential conflict of interest was reported by the authors.

References

- Agarwal, S., V.Y.S. Chen, and W. Zhang. 2016. “The information value of credit rating action reports: A textual analysis.” *Management Science*, 6: 2218–2240. <https://doi.org/10.1287/mnsc.2015.2243>.
- Ali, H., and A.F. Aysan. 2023. “What will ChatGPT revolutionize in financial industry?” SSRN. <http://dx.doi.org/10.2139/ssrn.4403372>.
- Ali, Z. 2023. “Economics of ChatGPT: A labor market view on the occupational impact of artificial intelligence.” SSRN. <http://dx.doi.org/10.2139/ssrn.4350925>.
- Alshater, M. 2022. “Exploring the role of artificial intelligence in enhancing academic performance: A case study of ChatGPT.” SSRN. <http://dx.doi.org/10.2139/ssrn.4312358>.
- An, J., W.Z. Ding, and C. Lin. 2023. “ChatGPT: Tackle the growing carbon footprint of generative AI.” *Nature*, 61: 586. <https://doi.org/10.1038/d41586-023-00843-2>.
- Beam, E.A. 2023. “Social media as a recruitment and data collection tool: Experimental evidence on the relative effectiveness of web surveys and chatbots.” *Journal of Development Economics*, 162: 103069. <https://doi.org/10.1016/j.jdeveco.2023.103069>.
- Beerbaum, D.O. 2023. “Generative Artificial Intelligence (GAI) ethics taxonomy—applying Chat GPT for robotic process automation (GAI-RPA) as business case.” SSRN. <http://dx.doi.org/10.2139/ssrn.4385025>.
- Beraja, M., D.Y. Yang, and N. Yuchtman. 2023. “Data-intensive innovation and the state: Evidence from AI firms in China.” *The Review of Economic Studies*, 90: 1701–1723, <https://doi.org/10.1093/restud/rdac056>.
- Buber, S., V. Chandrasekaran, R. Eldan, J. Henrke, E. Horvitz, E. Kamar, P. Lee, et al. 2023. “Sparks of artificial general intelligence: Early experiments with GPT-4.”

<https://doi.org/10.48550/arXiv.2303.12712>.

Brynjolfsson, E., D. Li, and L.R. Raymond. 2023. “Generative AI at work.” National Bureau of Economic Research Working Paper 31161. <http://www.nber.org/papers/w31161>.

ChatGPT, A. Zaremba, and E. Demir. 2023. “ChatGPT: Unlocking the future of NLP in finance.” SSRN. <http://dx.doi.org/10.2139/ssrn.4323643>.

Clarke, L. 2023. “Alarmed tech leaders call for AI research pause.” *Science*, 380: 120-121. <http://doi.org/10.1126/science.adi2220>.

Cao, Y., and J. Zhai. 2023. “Bridging the gap—the impact of ChatGPT on financial research.” *Journal of Chinese Economic and Business Studies*, 21: 1-15. <https://doi.org/10.1080/14765284.2023.2212434>.

Constantinescu, M., C. Vica, R. Uszkai, R., and C. Voinea. 2022. “Blame it on the AI? On the moral responsibility of artificial moral advisors.” *Philosophy & Technology*, 35. <https://doi.org/10.1007/s13347-022-00529-z>.

Debnath, R., R. Bardhan, D.U. Shah, K. Mohaddes, M.H. Ramage, R.M. Alvarez, and B.K. Sovacool. 2022. “Social media enables people-centric climate action in the hard-to-decarbonise building sector.” *Scientific Reports*, 12: 19017. <https://doi.org/10.1038/s41598-022-23624-9>.

Dis, E.A.M., J. Bollen, and W. Zuidema. 2023. ChatGPT: “Five priorities for research.” *Nature*, 614: 224-226. <https://doi.org/10.1038/d41586-023-00288-7>.

Dixit, R.K., A.K. Dixit, and R.S. Pindyck. 1994. *Investment under Uncertainty*. Princeton: Princeton University Press.

- 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>.
- Dwivedi, Y.K., N. Kshetri, L. Hughes, E.L. Slade, A. Jeyaraj, A.K. Kar, A.M. Baabdullah. 2023. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy." *International Journal of Information Management*, 71: 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>.
- Eckerli, F., and J. Osterrieder. 2021. "Generative adversarial networks in finance: an overview." arXiv preprint arXiv:2106.06364. <https://doi.org/10.48550/arXiv.2106.06364>.
- Eisfeldt, A.L., G. Schubert, and M.B. Zhang. 2023. Generative AI and firm values. National Bureau of Economic Research Working Paper 31222. <http://www.nber.org/papers/w31222>.
- Fischer, C., and R.G. Newell. 2008. "Environmental and technology policies for climate mitigation." *Journal of Environmental Economics and Management*, 55: 142–162. <https://doi.org/10.1016/j.jeem.2007.11.001>.
- George, A.S., A.S.H. George, and A.S.G. Martin. 2023. "A review of ChatGPT AI's impact on several business sectors." *Partners Universal International Innovation Journal*, 1: 9–23. <https://doi.org/10.5281/zenodo.7644359>.
- Gherhes, C., Z. Yu, T. Vorley, and L. Xue. 2023. "Technological trajectories as an outcome of the structure-agency interplay at the national level: Insights from emerging varieties of AI." *World Development*, 168: 106252. <https://doi.org/10.1016/j.worlddev.2023.106252>.

- Gil, D., S. Hobson, A. Mojsilovic, A. Mojsilovic, R. Puri, and J.R. Smith. 2018. "AI for management: An overview." In *The Future of Management in an AI World*, edited by J. Canals and F. Heukamp, 3-19. London: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-20680-2_1.
- Hamamoto, M. 2006. Environmental regulation and the productivity of Japanese manufacturing industries. *Resource and Energy Economics*, 28: 299–312. <https://doi.org/10.1016/j.reseneeco.2005.11.001>.
- Hansen, A.L., and S. Kazinnik. 2023. "Can ChatGPT decipher fedspeak?" SSRN. <http://dx.doi.org/10.2139/ssrn.4399406>.
- Hao, K. 2020. "OpenAI is giving Microsoft exclusive access to its GPT-3 language model." *MIT Technology Review*, September 23, 2020. <https://www.technologyreview.com/2020/09/23/1008729/openai-is-giving-microsoft-exclusive-access-to-its-gpt-3-language-model/>.
- Hassan, T.A., S. Hollander, L.V. Lent, and A. Tahoun. 2019. "Firm-level political risk: Measurement and effects." *The Quarterly Journal of Economics*, 134: 2135–2202. <https://doi.org/10.1093/qje/qjz021>.
- Herzenstein, M., S. Sonenshein, and U.M. Dholakia. 2011. "Tell me a good story and I may lend you money: The role of narratives in peer-to-peer lending decisions." *Journal of Marketing Research*, 48: S138–S149. <https://doi.org/10.1509/jmkr.48.SPL.S138>.
- Johnstone, N., I. Haščič, and D. Popp. 2010. "Renewable energy policies and technological innovation: Evidence based on patent counts." *Environmental and Resource Economics*, 45: 133–155. <https://doi.org/10.1007/s10640-009-9309-1>.

- Korangi, K., C. Mues, and C. Bravo. 2023. "A transformer-based model for default prediction in mid-cap corporate markets." *European Journal of Operational Research*, 308: 306-320. <https://doi.org/10.1016/j.ejor.2022.10.032>.
- Korinek, A. 2023. "Language models and cognitive automation for economic research." National Bureau of Economic Research Working Paper 30957. <https://doi.org/10.3386/w30957>.
- Korzynski, P., G. Mazurek, A. Altmann, J. Ejdyś, R. Kazlauskaitė, J. Paliszkievicz, K. Wach, and E. Ziemba. 2023. "Generative artificial intelligence as a new context for management theories: Analysis of ChatGPT." *Central European Management Journal*, 38: 3-13. <https://doi.org/10.1108/CEMJ-02-2023-0091>.
- Liebrenz, M., R. Schleifer, A. Buadze, D. Bhugra, and A. Smith. 2023. "Generating scholarly content with ChatGPT: Ethical challenges for medical publishing." *The Lancet Digital Health*, 5: e105-e106. [https://doi.org/10.1016/S2589-7500\(23\)00019-5](https://doi.org/10.1016/S2589-7500(23)00019-5).
- Lopez-Lira, A., and Y. Tang. 2023. "Can ChatGPT forecast stock price movements? Return predictability and large language models." SSRN. <http://dx.doi.org/10.2139/ssrn.4412788>.
- Lu, Q., Y. Luo, L. Zhu, M. Tang, X. Xu, and J. Whittle. 2023. "Developing responsible chatbots for financial services: A pattern-oriented responsible AI engineering approach." Unpublished manuscript. <https://arxiv.org/abs/2301.05517>
- Lu, Y., and Y. Zhou. 2021. "A review on the economics of artificial intelligence." *Journal of Economic Surveys*, 35: 1045-1072. <https://doi.org/10.1111/joes.12422>.
- Mann, S.P., B.D. Earp, S. Nyholm, J. Danaher, N. Moller, H. Bowman-Smart, J. Hatherley, et al. 2023. "Generative AI entails a credit-blame asymmetry." *Nature Machine Intelligence*, May 4, 2023. <https://doi.org/10.1038/s42256-023-00653-1>.

- Nabavi, E., and C. Browne. 2023. "Leverage zones in responsible AI: Towards a systems thinking conceptualization." *Humanities & Social Sciences Communications*, 10: 82. <https://doi.org/10.1057/s41599-023-01579-0>.
- Nguyen, J.H., and H.V. Phan. 2020. "Carbon risk and corporate capital structure." *Journal of Corporate Finance*, 64: 101713. <https://doi.org/10.1016/j.jcorpfin.2020.101713>.
- Noy, S., and W. Zhang. 2023. "Experimental evidence on the productivity effects of generative artificial intelligence." SSRN. <http://dx.doi.org/10.2139/ssrn.4375283>.
- OpenAI. 2023a. "ChatGPT plugins." <https://openai.com/blog/chatgpt-plugins>.
- OpenAI. 2023b. "Announcing OpenAI's bug bounty program." <https://openai.com/blog/bug-bounty-program#OpenAI>.
- Orzechowski, P., and J.H. Moore. 2022. "Generative and reproducible benchmarks for comprehensive evaluation of machine learning classifiers." *Science Advances*, 8: eabl4747. <https://doi.org/10.1126/sciadv.abl4747>.
- Quillian, L., J.J. Lee, and B. Honoré. 2020. "Racial discrimination in the US housing and mortgage lending markets: A quantitative review of trends, 1976–2016." *Race and Social Problems*, 12: 13-28. <https://doi.org/10.1007/s12552-019-09276-x>.
- Savage, N. 2023. "Synthetic data could be better than real data." *Nature*, April 27, 2023. <https://www.nature.com/articles/d41586-023-01445-8>.
- Scepanovic, S., M. Constantinides, D. Quercia, and S. Kim. 2023. "Quantifying the impact of positive stress on companies from online employee reviews." *Scientific Reports*, 13: 1603. <https://doi.org/10.1038/s41598-022-26796-6>.
- Seghier, M.L. 2023. "ChatGPT: Not all languages are equal." *Nature*, 615: 216. <https://doi.org/10.1038/d41586-023-00680-3>.

- Senior, M. 2023. "Precision financing." *Nature Biotechnology*, April 12, 2023. <https://doi.org/10.1038/d41587-023-00001-z>.
- Singh, H., and A. Singh. 2023. "ChatGPT: Systematic Review, Applications, and Agenda for Multidisciplinary Research." *Journal of Chinese Economic and Business Studies*, 21: 193-212. <https://doi.org/10.1080/14765284.2023.2210482>.
- Singh, U., A. Saraswat, H. Azad, K. Abhishek, and S. Shitharth. 2022. "Towards improving e-commerce customer review analysis for sentiment detection." *Scientific Reports*, 12: 21983. <https://doi.org/10.1038/s41598-022-26432-3>.
- Stevenson, M., C. Mues, and C. Bravo. 2021. "The value of text for small business default prediction: A deep learning approach." *European Journal of Operational Research*, 295: 758–771. <https://doi.org/10.1016/j.ejor.2021.03.008>.
- Thorp, H.H. 2023. "ChatGPT is fun, but not an author." *Science*, 379: 313. <https://doi.org/10.1126/science.adg7879>.
- Wan, X., J. Yang, S. Marinov, J.-P. Calliess, S. Zohren, and X. Dong. 2021. "An correlation in financial news networks and associated market movements." *Scientific Reports*, 11, 3062. <https://doi.org/10.1038/s41598-021-82338-6>.
- Wenzlaff, K., and S. Spaeth. 2022. "Smarter than Humans? Validating how OpenAI's ChatGPT model explains crowdfunding, alternative finance and community finance." SSRN. <http://dx.doi.org/10.2139/ssrn.4302443>.
- Yang, H. 2023. "How I use ChatGPT responsibly in my teaching." *Nature*, April 12, 2023. <https://doi.org/10.1038/d41586-023-01026-9>.
- Yang, K.C., and F. Menczer. 2023. "Large language models can rate news outlet credibility." Unpublished manuscript. <https://arxiv.org/abs/2304.00228v1>.
- Yin, K., F. Cai., and C. Huang. 2023. "How does artificial intelligence development affect green technology innovation in China? Evidence from dynamic panel data

analysis.” *Environmental Science and Pollution Research*, 30: 28066-28090.

<https://doi.org/10.1007/s11356-022-24088-0>.

Zhou, G., G. Chu, L. Li., and L. Meng. 2020. “The effect of artificial intelligence on

China’s labor market.” *China Economic Journal*, 13: 24-41.

<https://doi.org/10.1080/17538963.2019.1681201>.