



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Bridging the gap—the impact of ChatGPT on financial research

Citation for published version:

Cao, Y & Zhai, J 2023, 'Bridging the gap—the impact of ChatGPT on financial research', *Journal of Chinese Economic and Business Studies*, vol. 21, no. 2, pp. 177-191.
<https://doi.org/10.1080/14765284.2023.2212434>

Digital Object Identifier (DOI):

[10.1080/14765284.2023.2212434](https://doi.org/10.1080/14765284.2023.2212434)

Link:

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

Document Version:

Publisher's PDF, also known as Version of record

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.





Bridging the gap – the impact of ChatGPT on financial research

Yi Cao & Jia Zhai

To cite this article: Yi Cao & Jia Zhai (2023) Bridging the gap – the impact of ChatGPT on financial research, *Journal of Chinese Economic and Business Studies*, 21:2, 177-191, DOI: [10.1080/14765284.2023.2212434](https://doi.org/10.1080/14765284.2023.2212434)

To link to this article: <https://doi.org/10.1080/14765284.2023.2212434>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 13 May 2023.



Submit your article to this journal [↗](#)



Article views: 163



View related articles [↗](#)



View Crossmark data [↗](#)



Bridging the gap – the impact of ChatGPT on financial research

Yi Cao^a and Jia Zhai^b

^aManagement Science and Business Economics Group, Business School, University of Edinburgh, Edinburgh, UK; ^bBusiness School, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, PR China

ABSTRACT

In March 2023, the release of GPT-4 and its application, Copilot, astounds the world and thrusts AI into the spotlight in industry, and academia. The incredible superiority of GPT-4 is demonstrated by its ability to achieve high scores on almost all mainstream academic and professional standard exams, Copilot's capability to accomplish nearly all repetitive office work, and the rapid spread of its applications across massive areas of human society within weeks of its launch. These changes lead to the belief on the emergence of GPT-4 having a significant impact on academic research in the finance and accounting fields by establishing a consensus on the psychological acceptance of AI and rapidly eliminating technical barriers of using it. This paper presents practical examples to demonstrate GPT-4's effectiveness in sentiment analysis, ESG analysis, corporate culture analysis, and Federal Reserve opinion analysis, and provides instructive recommendations for applying it in these subject areas.



ARTICLE HISTORY

Received 21 April 2023

Accepted 1 May 2023

1 Introduction

February-March 2023 could be the craziest, most breathless month in the history of AI development to date in human society. We try to summarize what has happened in AI in a month that is both short and long. On February 13, Meta released Toolformer, telling the world that large language models (LLM) can already start utilizing tools, so that instead of solving all the problems themselves, such as computation, reasoning, etc., LLM can simply act as a strong interactive tools that translate human requirements into executable commands and utilize the most appropriate specialized software to complete the task. At the end of February, Meta released their LLM called LLaMA. After that, on 8 March, Visual ChatGPT has been released. It connects ChatGPT with a range of Visual Foundation models in order to use image as input and output in conversations. On 9 March, GigaGAN has been released that makes many people crazy on its capability of generating synthesized image by text inputs. On 13 March, Stanford Alpaca has been released as a fine-tuned LLaMA 7B model. On 14 March, GPT-4 has been released as the strongest LLM in this world. At the same day, Google

CONTACT Yi Cao  yi.cao@ed.ac.uk  Management Science and Business Economics Group, Business School, University of Edinburgh, 29 Buccleuch Place, Edinburgh EH8 9JS

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

released its generative AI API and Anthropic released their Claude LLM, which are considered as the biggest competitor of ChatGPT. At the same day, the LLM startup company Adapt.ai closes their Series-B round funding with \$350 million in funding. On 15 March, Midjourney released their 5-th generation models for text to image generation (they are the first model that draw human being's hands' details as real, which beat all previous text generation models). March 16 saw the craziest day in a month when Microsoft released a one-minute Copilot video showing how GPT enhanced office software can automate almost all repetitive office tasks, including making the presentation slides, drafting advertisement document, analysing financial data, writing budget report, drafting project proposal and more. This short one-minute video shows the world that (possibly) for the first time artificial intelligence could really replace human jobs, which causes widespread discussion.

While the GPT has made huge waves in the industry, there has also been a lot of sober commentary, such as Thorp (2023) questioning its creativity, Dwivedi et al. (2023) expressing concerns on its impact on education, and Lund et al. (2023) arguing a potential ethical problem in academia. Nonetheless, we remain optimistic that the emergence of GPT will not cause further problems for the future of academic research in finance and accounting, but is likely to be the biggest driver for a broader and deeper acceptance of AI and machine learning techniques in these subject areas. The emergence of GPT is likely to begin to build the consensus of acceptance for AI and machine learning techniques from a psychological perspective, and on the other hand, the emergence of GPT significantly lowers the technical barriers to apply the advanced machine learning techniques in finance and accounting researches.

This paper initiates with a discussion and analogy in Section 2 regarding the psychological consensus that GPT can achieve. In Section 3, we demonstrate how to use GPT to conduct textual analysis through four examples of machine learning applications in financial research: sentiment analysis, ESG analysis, corporate culture analysis, and analysis of Federal Reserve minutes. The conclusion of this paper is presented in Section 4.

2 Building the psychological consensus

Machine learning methods are presently extensively utilized in various domains of finance and accounting research. Nevertheless, there is still a degree of controversy surrounding their implementation in academic contributions, as well as a lack of consensus on researchers' psychological acceptance of these methods.

Considering the real academic contributions, prevalent comments are frequently made regarding machine learning, namely that while it may result in superior performance, such 'horse-racing' is not substantial for finance research. Additionally, the lack of interpretability associated with machine learning models, which renders them meaningless if their logic cannot be elucidated, is frequently cited. Undoubtedly, unexplained high performance and the occasional naive errors that occur are issues that cannot be tolerated in research, particularly in risk-sensitive domains such as self-driving cars, healthcare, and financial risk.

Nonetheless, a crucial issue that often goes unnoticed and impedes the implementation of machine learning is the psychological consensus of individuals. It is possible to draw an analogy between the present consensus on the acceptance of machine learning

and the renowned ‘copy machine study’ carried out by Ellen Langer at Harvard University in 1977 Langer, Blank, and Chanowitz (1978).

In the ‘copy machine study’ the researcher cuts into the line of individuals waiting for copies in the library with one of the following three questions to the next person in line (Langer, Blank, and Chanowitz 1978):

- (1) *(request only): ‘Excuse me, I have 5 pages. May I use the Xerox machine?’*
- (2) *(request with a real reason): ‘Excuse me, I have 5 pages. May I use the Xerox machine, because I’m in a rush?’*
- (3) *(request with a fake reason): ‘Excuse me, I have 5 pages. May I use the Xerox machine, because I have to make copies?’*

Obviously, the third reason is nonsensical; however, surprisingly, the experimental outcomes were favorable, and after consolidating all findings, the researchers found the following conclusions:

- (1) *60% of people let the researcher cut into the line*
- (2) *94% of people let the researcher cut into the line*
- (3) *93% of people let the researcher cut into the line*

The study achieved notability for revealing one of the most influential words that influences human behavior: ‘because’. Langer, Blank, and Chanowitz (1978)’s research demonstrated that individuals are likely to adopt a particular behavior as long as they can rationalize it in their mind (‘I do it because. . .’), even if the justifications are not particularly nonsensical. This psychological phenomenon suggests that people simply require a reason or excuse for their actions when confronted with unfamiliar behaviors or novel concepts. When individuals universally accept a particular ‘reason’ for justification, an implicit consensus of acceptance regarding that behavior is established. It is apparent that such a consensus has yet to be widely established in the research areas of machine learning in finance. Nevertheless, the establishment of this consensus may begin with GPT-4, as indicated by the following three primary reasons.

The first reason for the potential establishment of a subconscious consensus of acceptance on machine learning is the exceptional performance of GPT-4 in various professional and academic tests, which is difficult for humans to perform OpenAI (2023). This consistent exceptional outperformance across multiple domains has increased people’s trust in GPT-4, as well as machine learning in general. Although machine learning does not possess intelligence, many people have started to believe otherwise due to this trust in GPT’s capabilities. For instance, GPT-4 passed the mock version of the Uniform Bar Examination, which is one of the most challenging professional exams in the United States, scoring in the top 10% of candidates (Table 1 and Figure 4 in OpenAI (2023)). Furthermore, it achieved top scores on various U.S. national university entry and post-graduate entry exams, including the SAT Reading & Writing (top 7%), SAT Math (top 11%), and GRE exam (top 22% on average). Additionally, GPT-4 excelled in university exams in English language, economics, physics, history, and other subjects, ranking around the top 20–10%. In other words, the competencies that GPT-4 has achieved from the ‘learning’ process have

surpassed those of at least 80% of humans who have completed these standardised exams. However, the victories of AlphaGo over Go champion Ke Jie in 2016 Silver et al. (2016) and AlphaFold's distant lead in protein structure prediction Jumper et al. (2021) are equally remarkable, they are not as relatable to ordinary people's lives. Thus, people may find it challenging to embrace such achievements. In sum, in contrast to other machine learning accomplishments, GPT-4's exceptional performance in numerous standardized tests has provided people with a compelling reason to accept this technology. As previously mentioned, individuals tend to require a justification or excuse to embrace new concepts or innovations. GPT-4's exemplary performance in standard exams that individuals are familiar with, such as the Uniform Bar Examination, the SATs, the GRE, and university exams, provides the reasoning and justification needed for individuals to embrace this new technology. Even if individuals are hesitant to trust machine learning, they may have no choice but to accept it once confronted with GPT-4's exceptional examination results that they may not be able to replicate themselves.

The second reason is the potential for a new smart-industrial revolution that cannot be stopped. As previously discussed, people often need a reason or excuse to accept something new. The release of GPT-4 has provided such a reason through its outstanding performance in a variety of standardised examinations. In addition, the first large-scale application that demonstrated GPT-4's capabilities was <https://www.youtube.com/watch?v=S7xTBa93TX8> Microsoft 365 Copilot, which has already caused a sensation around the world Spataro (2023). The one-minute introductory video gives people a glimpse of the revolution that machine learning may bring to the industry. Copilot, an enhanced version of Microsoft Office powered by GPT-4, can automatically perform repetitive and routine tasks in daily office work, such as email responses, presentation slide preparation, budget drafting, and project proposal writing. Since these tasks make up a significant portion of many people's daily work, discussions have arisen about the possibility of jobs being replaced by machine learning models Mok and Zinkula (2023); Dwivedi et al. (2023). Although people may initially distrust these models, the fear of losing their jobs may be the catalyst for acceptance, just as the first commercial railway between Liverpool and Manchester was built in 1830, more than half a century after James Watt's improvement of the steam engine in 1776, during which time many locomotives, such as Richard Trevithick's Puffing Devil in 1801 Trevithick (1872) and George Stephenson's Blucher in 1814 Smiles (1859), were widely questioned. In the current era, GPT-4 can be compared to Stephenson's Blucher locomotive in 1814. It was only 16 years after the improved locomotive that the first commercial railway was constructed in 1830, heralding the most significant efficiency gains of the first industrial revolution. Similarly, within a time frame certainly shorter than 16 years from the commencement of GPT-4, it is plausible that we will witness another industrial revolution, wherein machine learning and artificial intelligence are extensively applied to daily work, resulting in a substantial leap in productivity.

The third reason is that widespread commercial applications of GPT-4 begin immediately after its release. Stephenson's Blucher was only one of the key drivers of the Industrial Revolution; it was the mass application of the steam engine in manufacturing across many industries throughout the entire society that was the fundamental factor. The beginning of the GPT-4's large-scale commercial application nowadays is the equivalent of the steam engine in the Industrial Revolution of 1814.

On 1 March 2023, OpenAI announced the opening of the ChatGPT API, powered by the large natural language model 'GPT 3.5 Turbo', with an open API fee model of \$0.002/1000 tokens, a 90% reduction in price from the previous generation. In addition to this, OpenAI has also opened up the Whisper API, a speech to text model that can transcribe or translate speech. It supports dozens of languages, including English, Chinese, Arabic, Japanese, German, Spanish and more, and costs \$0.006 per minute to use the Whisper API. Microsoft's search engine, Bing, combines the technology of GPT4 to combine search, creation and chat in a multi-faceted way to meet the individual needs of users; Microsoft's enterprise services platform Dynamic 365, a SaaS service that incorporates CRM, ERP and other functions, has combined ChatGPT's technology to launch Dynamics 365 Copilot, which provides people across businesses from sales, service and marketing to supply chain. Microsoft's developer tools, Power Platform, in conjunction with ChatGPT technology, launched with Power Virtual Agents and AI Builder, which provide a low code approach and developer-oriented generated AI capabilities.

In addition to applications conducted by OpenAI's parent company Microsoft, a very large number of startups have quickly followed by integrating GPT technology to enhance their products. Jasper, intuitive AI copywriter, has a number of products that access GPT-3 for commercial copywriting, including Jasper Starter for abstracts, product descriptions and blog posts using predefined templates, and Boss Mode, a writing assistant that enables flexible and easy business copywriting. By using Boss Mode, users only need to edit 20% of the content themselves. Jasper for Business is a dedicated service for B-side copywriting, with the ability to generate customised copy. Users can create new custom workflows and develop writing assistants specifically tailored to their needs; Jasper Art can generate images for advertising and blog posts. On 7 March 2023, Brex, a financial technology platform in U.S., announced the launch of an advanced AI tool based on OpenAI GPT-4 technology. Through the Brex Empower platform, customers can use natural language to describe requests in a chat interface to obtain the appropriate financial-related information or more complex financial analysis. Slack is a team collaboration tool owned by Salesforce, similar to Teams. On 7 March 2023, Slack announced the introduction of OpenAI's ChatGPT, which allows users to ask ChatGPT questions directly in the app to get answers to their questions. HireStack is an HR company that uses GPT-3 to generate interview questions and their corresponding answers. Pencil Advertising, a creative marketing and design agency, uses GPT-3 to generate effective, creative and brand-appropriate advertising copy and video. It enables creative marketing around the product. Snapchat, Discord and Quora have all launched chatbots My AI, Clyde and Poe respectively, based on the ChatGPT API. Quizlet, a global learning platform, launched on 28 February 2023 Q-Chat, an AI tutor based on ChatGPT, which asks adaptive questions based on relevant learning materials and generates interesting chat content to interact with students. A US-based fresh grocery e-commerce platform, Instacart, uses ChatGPT, combined with product information from over 75,000 retail partner shops on the platform, to help customers achieve their open shopping goals. The consumer app Shopify, through its integration with ChatGPT, will scan millions of items to quickly find what buyers are looking for and provide personalised recommendations based on the user's textual description of their needs. The data analysis software Anania, integrated with ChatGPT, allows the user's data to be analysed through simple, natural English conversation. The AI-assisted programming platform, Tabnine, incorporates GPT-4 to predict and

infer the user's next line of code based on context and syntax, or the user types a comment and Tabnine suggests the correct code for them. Lyrics generation tool Jarvis, based on GPT-3, can generate lyric suggestions based on any artist, genre, title and lyric prompt. The decision assistant software Rationale, based on GPT-3, lists the pros and cons of the decision based on contextual learning algorithms, generates a SWOT analysis or conducts a multi-criteria analysis to help users weigh up their options. These examples are just the tip of the iceberg of startups using GPT to enhance their products that have sprung up in the month since ChatGPT was officially released. These widespread applications in almost every aspect of life have contributed significantly to the consensus on the acceptance of GPT.

Thus, the three reasons mentioned above, i.e. the excellent results in almost all the most popular academic and professional exams, the revolution in smart technology triggered by the GPT itself, and the massive smart applications based on the GPT that are in every aspect of life, are the key reasons for the acceptance of machine learning (and even AI) techniques and the formation of a consensus of the widespread acceptance.

3 Smoothing the technological barrier

For academic researchers in the fields of finance and accounting, machine learning can help them in many areas to solve otherwise difficult research questions. But applying machine learning models is often very difficult for those without a professional computer programming background, and GPT-4 is probably the most powerful tool to date, providing a way to accomplish the task with interaction in a multilingual environment using ordinary conversation, removing technical barriers for their in-depth research topics.

However, GPT-4 is not a 'Harry Potter magic wand' in the hands of researchers that can turn everything it touches into gold, allowing them to complete complex research tasks quickly. However, it is more like a 'Harry Potter flying broomstick' that can magically enhance the convenience, efficiency, and accuracy for researchers utilizing machine learning techniques. The key lies in how to master the ability to control the 'broomstick' to soar over towering mountains and rugged peaks, rather than being thrown to the ground by the broomstick. This is a new course for researchers to explore.

Prompts are essential to guide the outcome of the GPT-4 for a particular task. A prompt can be a short paragraph that provides context and instructions for the GPT-4 to generate results, or it can be a complex description, depending on the difficulty of the task. In this section, we show that appropriate prompts in GPT-4 lead to answers to several common research questions in finance and accounting, which often require in-depth application of machine learning models. These questions include sentiment analysis, ESG keyword analysis and sentiment, company culture, and Federal Reserve Board Press.

3.1 Sentiment analysis

In finance, sentiment values extracted from textual data is termed as alternative data in the last decade and has shown many meaningful applications in financial empirical and even quantitative trading. However, extracting sentiment values from textual information (e.g. news, reports, social media) is technically challenging and generally requires training

a neural network model using a labelled dataset (Such data sets are typically read by humans who then determine the sentiment value of the text and label the text as 'optimistic', 'pessimistic', 'neutral', etc.) and then using the trained model to make judgements about the sentiment value of the textual information.

We can now easily to use GPT-4 to quantify the sentiment value of a piece of text. We use the following prompt:

On a scale of -1 to 1, where -1 is the most negative and 1 is the most positive, what is the sentiment value of the following text: _Text_ Requiring_Sentiment_Analysis_

We use a few news headlines from the Markets section of Financial Times on 18 April 2023 as an example to empirically analyse the sentiment values generated by GPT-4, as shown in [Figure 1](#). From these five examples we can see that the GPT-4 provides accurate sentiment values, with not only scores but also the analysed content of the text, increasing the confidence of the scores.

- (1) Sentiment of *'HSBC urged to break up by largest shareholder and accused of "exaggerating" risk'*
- (2) Sentiment of *'EU faces last-ditch challenge from exchanges over trading reforms: Bourses warn consolidated tape plan could harm smaller European markets'*
- (3) Sentiment of *'Pension fund CalSTRS braced for writedowns in 50bnpropertyportfolio : Warningfrom 306bn fund is latest sign that higher interest rates have upended US commercial real estate market'*
- (4) Sentiment of *'Changes to UK financial regulation will not weaken it, but make it more effective: The government's "Edinburgh reforms" are designed to help the sector drive growth'*
- (5) Sentiment of *'BlackRock avoids 60/40 portfolio despite stock and bond rebound: Traditional mix of stocks and bonds is ill-suited to world of high inflation, says world's biggest asset manager'*

3.2 ESG analysis

When we analyse the ESG content contained in textual information, we usually employ two types of analysis, firstly the percentage of ESG keywords and secondly the sentiment of the text regarding ESG, i.e. the ESG sentiment.

We found that GPT-4 works very well in the generation of the required keyword list. We use the following prompt to generate 20 ESG keywords (we can adjust the number of keywords to any number we want, e.g. 50, 100 etc.) as shown in [Figure 2](#).

Generate a list of _Number_Of_Keywords_ keywords most relevant to ESG

The generated results are shown in [Figure 2](#). When we adjust the number of keywords required, we get the list of the required length.

Afterwards, we use a sustainable report released by Alphabet in 2022 to evaluate the ESG keywords percentage and ESG sentiment. The report we use is <https://sustainability.google/reports/2022> Environmental Report. We randomly copy some paragraph in the

Based on the text provided, I would say the sentiment value is around -0.7, indicating a relatively negative sentiment. The text contains negative language, such as 'urged to break up,' 'accused of exaggerating risk,' and the implication of conflict between the largest shareholder and the company.

(1) Sentiment of “*HSBC urged to break up by largest shareholder and accused of ‘exaggerating’ risk*”

The sentiment value of the following text is slightly negative, around -0.2 to -0.3. The article talks about a last-ditch challenge from exchanges over trading reforms, with bourses warning that the consolidated tape plan could harm smaller European markets. The use of phrases such as "last-ditch challenge," "warn," and "harm" imply a negative tone. However, the sentiment is not extremely negative, as the article also notes that this challenge is a result of the EU's push for trading reforms, which suggests a potential positive outcome.

(2) Sentiment of “*EU faces last-ditch challenge from exchanges over trading reforms: Bourses warn consolidated tape plan could harm smaller European markets*”

The sentiment value of the following text would be around -0.7 to -0.8, which indicates a very negative sentiment. The text talks about the warning from Calstrs, a large pension fund, about the potential writedowns in its \$50bn property portfolio due to the impact of higher interest rates on the US commercial real estate market. This news is likely to be perceived negatively by investors and market participants, hence the negative sentiment score.

(3) Sentiment of “*Pension fund Calstrs braced for writedowns in 50bnpropertyport folio : Warningfrom306bn fund is latest sign that higher interest rates have upended US commercial real estate market*”

The sentiment value of the following text is 0.5, which is very positive. The text suggests that changes to UK financial regulation will not weaken it, but rather make it more effective. The government's "Edinburgh reforms" are designed to help the sector drive growth, which implies a positive outlook for the UK financial sector.

(4) Sentiment of “*Changes to UK financial regulation will not weaken it, but make it more effective: The government’ s ‘Edinburgh reforms’ are designed to help the sector drive growth*”

The sentiment value of the following text is 0.0. The text presents a neutral statement about BlackRock’s investment strategy and does not express any particularly positive or negative sentiment.

(5) Sentiment of “*BlackRock avoids 60/40 portfolio despite stock and bond rebound: Traditional mix of stocks and bonds is ill-suited to world of high inflation, says world’ s biggest asset manager*”

Figure 1. The GPT-4 generated sentiment values of news headlines published in Financial Time on 18 April 2023. The values are between -1 to $+1$, where -1 denotes the most negative and $+1$ is the most positive.

Section ‘Performance highlights’ on page 7 and use the following prompt to generate the ESG keywords percentage:

*What is the percentage of the above generated ESG keywords in the following paragraph:
_Text_To_Analyse_*

The **Figure 3** shows the GPT-4 generated percentage based on the keywords that have been generated by GPT-4 itself. In addition, we use the following prompt to generate the ESG sentiment:



Figure 2. The GPT-4 generated 20 most important ESG keywords.

*On a scale of -1 to 1, where -1 is very negative and 1 is very positive, what is the sentiment value regarding to the above-generated ESG keyword list for the following text:
_Text_To_Analyse_*

The generated ESG sentiment value is shown in [Figure 4](#). From the results, it appears that GPT-4 knows how to analyse text and generate keyword-related sentiment values based on the list of ESG keywords it has just generated itself.

3.3 Corporate culture

Corporate culture has received a lot of attention in recent years in the field of corporate finance Li et al. (2021,a). We try to use GPT-4 to do certain analysis of

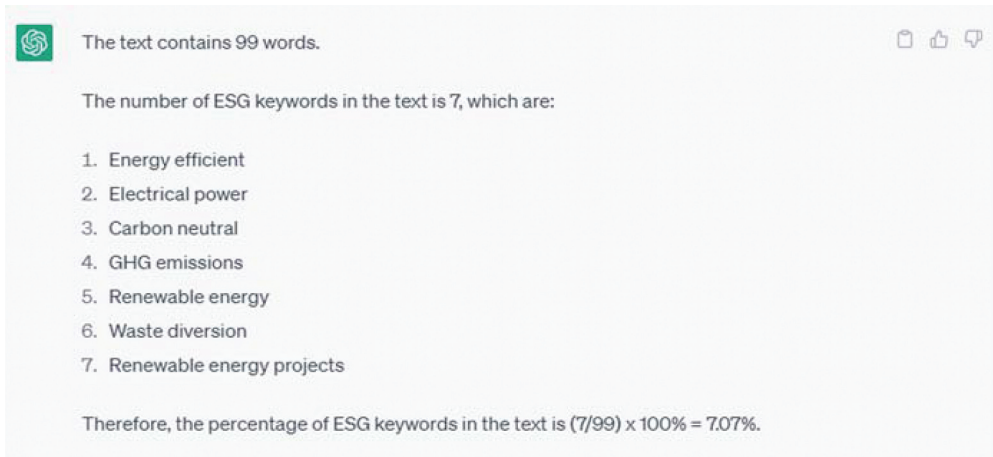


Figure 3. The GPT-4 generated keywords percentage based on the generated 20 most important ESG keywords.

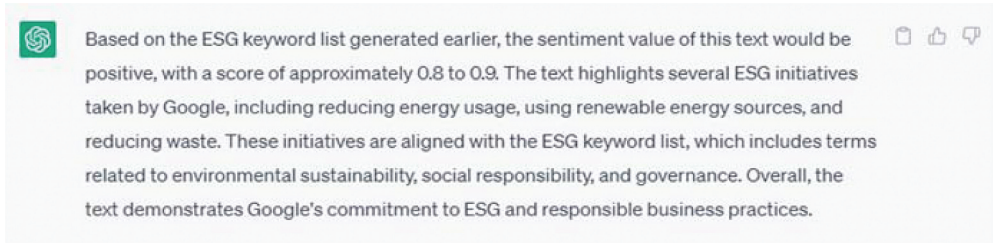


Figure 4. The GPT-4 generated ESG sentiment value based on the earlier generated 20 most important ESG keywords.

corporate culture. Firstly, we can use the following prompt to generate a list of keywords relating to our corporate culture, and we can control the number of keywords we need:

Generate a list of the _Number_Of_Keywords_ keywords most relevant to a company's innovation

Alternatively, we can also use a list of keywords of corporate culture summarised in existing literature, e.g. Table IA3 of Li et al. (2021). We use the following simple prompt to 'teach' GPT-4 the list (using the Panel A of the table as the example):

We have an innovation keyword list as follows: _Keywords_List_

After feeding the keywords to GPT-4, we conduct two types of analysis on corporate culture, the first is the percentage of the keywords and the second is the corporate culture sentiment. We randomly select an earning call transcript, released by the corporate <https://seekingalpha.com/article/4594611-midwest-energy-emissions-corp-meec->

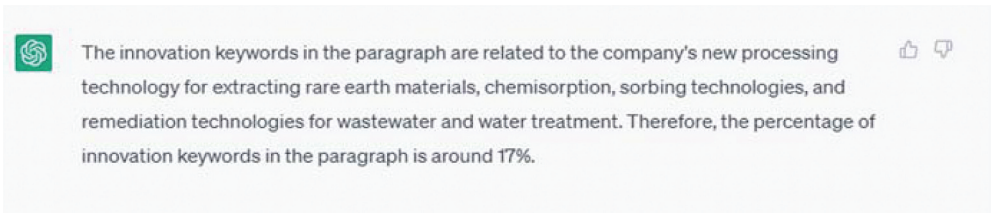


Figure 5. The GPT-4 generated keywords percentage for the earning call transcript of <https://seekingalpha.com/article/4594611-midwest-energy-emissions-corp-meec-q4-2022-earnings-call-transcript> "Midwest Energy Emissions Corp" for Q4 2022 based on the "learned" corporate culture innovation keywords in Table IA3 in Li et al. (2021).



Figure 6. The GPT-4 generated corporate culture sentiment value of the earning call transcript of <https://seekingalpha.com/article/4594611-midwest-energy-emissions-corp-meec-q4-2022-earnings-call-transcript> "Midwest Energy Emissions Corp" for Q4 2022 based on the earlier "learned" corporate culture innovation keywords in Table IA3 in Li et al. (2021).

q4-2022-earnings-call-transcript"Midwest Energy Emissions Corp" for Q4 2022. We first use the following prompt for the percentage of the keywords:

What is the percentage of the innovation keywords in the following paragraph: _Text_

We use full transcript to generate the corporate culture keyword percentage in Figure 5. In addition, we use the following prompt to generate the corporate culture sentiment:


On a scale of -1 to 1, where -1 is very negative and 1 is very positive, what is the sentiment value of the company culture based on the above keywords of the following text: _Text_

Figure 6 shows the corporate culture sentiment value based on the 'learned' keyword list.

3.4 Federal reserve board press

The attitude of the Federal Reserve's Federal Open Market Committee (FOMC) meeting is often considered to be one of the strongest drivers in financial markets change, and whether the Fed speaks dovishly or hawkishly is often considered to be decisive in guiding markets directions.


We tried to use GPT-4 to determine the dovish or hawkish attitude of the content of the Fed FOMC minutes. Using the following prompt, we try to generate a numerical value

 The provided text indicates a relatively dovish stance by the Federal Reserve, with a focus on financial conditions and inflation expectations. The following attitude values can be assigned on the -1 to 1 scale:

- Financial conditions tightening: -0.5
- Concern about bank lending pullback: -0.5
- Underperformance of large U.S. banks: -0.5
- Market-based measures of inflation compensation rising: 0.3
- Forward inflation compensation indicating well-anchored expectations: 0.5
- Inflation expectations from surveys increasing at shorter horizons: 0.2
- Inflation expectations from surveys little changed at longer horizons: 0.1

Overall, the text suggests a moderately dovish stance, with some concerns about the impact of recent developments in the banking sector on financial conditions, but a generally positive outlook for inflation expectations.

The value of Dovish or hawkish views of the first two paragraphs under the section of “Developments in Financial Markets and Open Market Operations” at page 3 of Minutes of the FOMC on March 21-22, 2023

 Based on the language used in the statement, it's difficult to determine a precise attitude value for the Fed's talks. However, we can make some inferences based on the information provided.

Regarding the labor market, the Fed notes that conditions remained tight, with robust job gains and the unemployment rate near a historical low. This suggests a relatively hawkish attitude, as the strong labor market could lead to inflationary pressures. Similarly, the statement notes that consumer price inflation remained elevated early in the year, with total PCE price inflation at 5.4% over the 12 months ending in January, and core PCE price inflation at 4.7%. This also suggests a hawkish attitude, as elevated inflation could prompt the Fed to tighten monetary policy.

On the other hand, the statement notes that real GDP growth seemed to be expanding at a modest pace in the first quarter, and that growth in business fixed investment was slowing and residential investment was continuing to decline. This suggests a more dovish attitude, as slower economic growth could prompt the Fed to maintain its accommodative monetary policy stance. Additionally, the statement notes that some indicators pointed to softening factory output in the coming months, and that many foreign central banks continued their monetary policy tightening, although some have either paused or indicated that a pause soon was possible. These factors also suggest a more dovish attitude.

Overall, based on the information provided, it's difficult to give a precise attitude value for the Fed's talks. However, the language used in the statement suggests a somewhat mixed attitude, with some hawkish and some dovish elements. A possible attitude value for this statement could be around 0.2, indicating a slightly hawkish bias.

The value of Dovish or hawkish views of the first three paragraphs under the section of “Staff Review of the Financial Situation” at page 3-4 of Minutes of the FOMC on March 21-22, 2023

Figure 7. The GPT-4 generated value of Dovish or hawkish views of <https://www.federalreserve.gov/newsevents/pressreleases/monetary20230412a.htm> Minutes of the FOMC on March 21–22, 2023.

between -1 and $+1$ representing the opinion value of FOMC, where -1 is the most dovish and $+1$ is the most hawkish:

On a scale of -1 to 1 , where -1 is very dovish and 1 is very hawkish, what are the attitude values for the following Fed talks: _FOMC_Text_

We use the <https://www.federalreserve.gov/newsevents/pressreleases/monetary20230412a.htmMinutes> of the FOMC on March 21–22, 2023 as an example. We copy part of the text on page 3–4 to test GPT-4 and receive interesting results as [Figure 7](#). From the results, the GPT-4 can analyse the information in the minutes very accurately and give a very logical and reasonable conclusion of doves or hawks. However, sometimes it is not possible to generate a single value of opinion, but rather different values corresponding to numerous aspects, as in the upper part of [Figure 7](#). In the second half of [Figure 7](#), the GPT-4 eventually generates a single value for the view, which is very trustworthy in terms of the logic of the analysis. The capability of GPT-4 in the analysis of Fed talk is also addressed in Hansen and Kazinnik (2023).

The value of Dovish or hawkish views of the first two paragraphs under the section of ‘Developments in Financial Markets and Open Market Operations’ at page 3 of <https://www.federalreserve.gov/newsevents/pressreleases/monetary20230412a.htmMinutes> of the FOMC on March 21–22, 2023

The value of Dovish or hawkish views of the first three paragraphs under the section of ‘Staff Review of the Financial Situation’ at page 3–4 of <https://www.federalreserve.gov/newsevents/pressreleases/monetary20230412a.htmMinutes> of the FOMC on March 21–22, 2023

In this section, we employ four distinct scenarios to showcase the potential advantages of GPT-4 in the domains of finance and accounting, specifically in removing technical obstacles to the application of machine learning models to a certain extent. The advent of GPT-4 enables researchers to utilize cutting-edge natural language processing models through conversations on a range of prevalent topics, with almost zero requirement of extensive training in Python programming.

4 Conclusion

This paper discusses the implications and impact of the emergence of GPT-4 on the field of finance and accounting. The emergence of GPT-4 is considered to be a turning point for researchers in these fields to accept machine learning techniques fully. Prior to this, although some researchers had started to apply various machine learning models, many voices of skepticism were raised in mainstream finance area, which included concerns about the interpretability of models, black boxes, the trustworthiness of models, and the theoretical contribution of the high accuracy that models bring to financial theory. The emergence of GPT-4 is likely to lead to a slow but comprehensive acceptance of machine learning models for three reasons: (1) the outstanding performance of these models on various standardized tests, (2) their potential to replace repetitive human work, and (3) their widespread commercial application that is penetrating almost all areas of society.

This study attempts to use GPT-4 to answer four common types of questions in the application of machine learning models in finance and accounting, namely sentiment analysis, ESG content analysis, corporate culture analysis, and Fed opinion analysis. By employing various prompts, this research analyses news headlines from the Financial Times, Google’s sustainability report, the earnings call transcript of Midwest Energy Emissions Corp, and Fed FOMC meeting minutes. The findings demonstrate that GPT-4

is highly accurate and efficient in generating quantitative and logical analyses of the textual content.

This paper does not present additional empirical financial analyses of the results generated by GPT-4. Rather, it offers specific examples and recommendations for using GPT-4 as a potential technical aid. The authors believe that the emergence of GPT-4 has the potential to significantly reduce the technical barriers that finance and accounting researchers face when applying state-of-the-art natural language processing models. With minimal technical effort in GPT-4, researchers can focus on the research and analysis of domain-specific problems in finance and accounting. This shift would likely lead to a surge in research advancement within the field.

Acknowledgments

This work was supported by the National Natural Science Foundation of China, Young Scientists Fund [grant number 72101207].

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The work was supported by the National Natural Science Foundation of China, Young Scientists Fund [72101207].

Notes on contributors

Yi Cao is the senior lecturer in management science at University of Edinburgh business school. His research interest lies in the intersection of machine learning and financial topics, i.e., asset pricing, option pricing, risk assessment and trading behaviours. He published in *European Journal of Operational Research*, *IEEE Transactions on Neural Networks and Learning System*, *Decision Support Systems*, and *Quantitative Finance*.

Jia Zhai is the Senior Associate Professor at XJTLU International Business School. Her research is mainly in the areas of financial technology and applied financial econometrics, e.g. machine learning in finance, option pricing, sentiment analysis, and volatility forecasting. She published in *European Journal of Operational Research*, *Decision Support Systems*, *Quantitative Finance*, *Neurocomputing*, and *Expert Systems with Applications*.

References

- Dwivedi, Y. K., N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, A. K. Kar, A. M. Baabdullah, et al. 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. doi:10.1016/j.ijinfomgt.2023.102642.
- Hansen, A. L., and S. Kazinnik. "Can ChatGpt Decipher Fedspeak?." Available at SSRN (2023).
- Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, et al. 2021. "Highly Accurate Protein Structure Prediction with Alphafold." *Nature* 596 (7873): 583–589. doi:10.1038/s41586-021-03819-2.

- Langer, E. J., A. Blank, and B. Chanowitz. 1978. "The Mindlessness of Ostensibly Thoughtful Action: The Role of "Placebic" Information in Interpersonal Interaction." *Journal of Personality & Social Psychology* 36 (6): 635. doi:10.1037/0022-3514.36.6.635.
- Li, K., F. Mai, R. Shen, X. Yan, and I. Goldstein. 2021. "Measuring Corporate Culture Using Machine Learning." *The Review of Financial Studies* 34 (7): 3265–3315. doi:10.1093/rfs/hhaa079.
- Lund, B. D., T. Wang, N. R. Mannuru, B. Nie, S. Shimray, and Z. Wang. 2023. "Chatgpt and a New Academic Reality: Artificial Intelligence-Written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing." *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.24750.
- Mok, A., and J. Zinkula, 2023. ChatGpt May Be Coming for Our Jobs. Here are the 10 Roles That AI is Most Likely to Replace. <https://www.businessinsider.com/chatgpt-jobs-at-risk-replacement-artificial-intelligence-ai-labor-trends-2023-02?r=USIR=T>. [Online; accessed 16-April-2023].
- OpenAI, 2023. Gpt-4 technical report. arXiv:2303.08774.
- Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, et al. 2016. "Mastering the Game of Go with Deep Neural Networks and Tree Search." *Nature* 529 (7587): 484–489. doi:10.1038/nature16961.
- Smiles, S. 1859. *The Life of George Stephenson, Railway Engineer*. Boston: Ticknor and Fields.
- Spataro, J., 2023. Introducing Microsoft 365 Copilot - Your Copilot for Work. <https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/>. [Online; accessed 16-April-2023].
- Thorp, H. H. 2023. "Chatgpt is Fun, but Not an Author." *Science* 379 (6630): 313. doi:10.1126/science.adg7879.
- Trevithick, F. 1872. *Life of Richard Trevithick: With an Account of His Inventions*. Vol. 1. London; New York: E. & FN Spon.