Finance Research Letters 59 (2024) 104786

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Can ChatGPT assist in picking stocks?

Matthias Pelster *, Joel Val

Mercator School of Management, Duisburg–Essen University & european center for financial services (ecfs), Lotharstraße 65, 47057, Duisburg, Germany

ARTICLE INFO

JEL classification:
G12
G14
G17

Keywords:
Information processing
Artificial intelligence (AI)
ChatGPT

ABSTRACT

This paper studies whether ChatGPT-4 with access to the internet is able to provide valuable investment advice and evaluate financial information in a timely manner. Using a live experiment, we find a positive correlation between ChatGPT-4 ratings and future earnings announcements and stock returns. We find evidence that ChatGPT-4 adjusts ratings in response to earnings surprises and news events information in a timely manner. An investment strategy based on "attractiveness ratings" by ChatGPT-4 yields positive returns.

1. Introduction

In financial markets, news are of pivotal importance. News information provide the necessary data for investors to determine the intrinsic value of assets and make sound investment decisions. However, the sheer volume of (possibly conflicting) information in today's information age may be too much for investors to evaluate thoroughly (Kahneman, 1973; Hirshleifer and Teoh, 2003). This is especially true for individual investors who, compared to institutional investors, often face greater constraints in terms of time and technology (Barber and Odean, 2008).

ChatGPT, an advanced artificial intelligence (AI) language model, emerges as a potentially beneficial tool for investors. By leveraging its natural language processing (NLP) capabilities, ChatGPT can potentially exploit patterns observed in the past to distill large volumes of information into concise summaries, thereby aiding investors in efficiently sifting through the plethora of news and data. ChatGPT can provide responses to specific queries and facilitate an interactive and tailored exchange that aligns with investors' needs. As the technology is available for all types of investors, it has the potential to lower information costs and to democratize financial knowledge and increase the capacity of individual investors to formulate well-informed investment decisions.

In this study, we evaluate ChatGPT's ability to synthesize financial news and explore its accuracy in determining actionable investment insights. We conduct a live experiment around the 2023-Q2 earnings announcement period to test whether ChatGPT reliably evaluates earnings surprises and other news events. Earnings announcements are a natural choice for our tests. They are scheduled major corporate events and create information-rich environments that harbor complexity and subtlety which is often challenging for investors to evaluate (Noh and Zhou, 2022; Rennekamp et al., 2022; Allee et al., 2021). Using a live experiment is important to avoid biased results. Specifically, it is impossible to analyze the performance of ChatGPT using historical data as it is not possible to make sure that ChatGPT does not use "future" data to make recommendations in such instances.

Our results indicate that earnings forecasts issued by ChatGPT are positively correlated to actual earnings, after we control for consensus forecasts. "Attractiveness" ratings positively correlate to future stock returns. Moreover, we find evidence that suggests that ChatGPT reacts to news information in a timely manner.

* Corresponding author.
E-mail address: matthias.pelster@uni-due.de (M. Pelster).

https://doi.org/10.1016/j.frl.2023.104786
Received 15 October 2023; Received in revised form 13 November 2023; Accepted 21 November 2023
Available online 23 November 2023

Studying earnings announcement forecasts and stock attractiveness ratings, we explore two use cases that lend themselves to investors employing ChatGPT as a supporting tool in their decision-making processes. The literature finds that positive (negative) surprises result in positive (negative) cumulative abnormal returns (CARs) over several months after the announcement (Ali et al., 2020; Doyle et al., 2006; Milian, 2015). Investors may ask ChatGPT to assist in the selection of stocks to invest prior to earnings announcements. This holds in the more general case of stock attractiveness ratings as well. There are cognitive—and temporal—limits to how much information investors can process. It is far easier, for example, to thoroughly consider an investment set of pre-selected alternatives than the complete universe of stocks.

We make an important contribution to the literature on ChatGPT's use-cases in finance (e.g., Dowling and Lucey, 2023; Fieberg et al., 2023; Niszczota and Abbas, 2023). By providing tutoring-like support and helping to interpret data in a conversational manner, ChatGPT and similar technologies are likely to affect the decision-making processes of investors. The rapid integration of ChatGPT and similar technologies in various industries suggests that, in the future, people will more rely on information that has been retrieved, summarized, and analyzed by AI. Financial institutions are already working on implementations that assist customers and employees in extracting and interpreting information (Bautzer and Nguyen, 2023). Our findings suggest that the technology may already be helpful in that regard.

2. What is ChatGPT?

ChatGPT, based on OpenAI's GPT-4 architecture, is a state-of-the-art NLP designed to generate human-like text based on given prompts. The model is built upon the transformer architecture introduced by Vaswani et al. (2017) and leverages a deep learning approach with several layers of attention mechanisms to process and generate language. Due to its vast training data and advanced architecture, ChatGPT excels at multiple applications such as content generation, translation, and even complex tasks like code writing (Brown et al., 2020).

In Finance, large language models (LLMs) can be used in automated customer support, financial analysis, fraud detection, algorithmic trading, personal finance management, regulatory compliance, education, or consulting (Cucchiara, 2023). Given the wide range of applications, it is not surprising that Morgan Stanley is already employing ChatGPT in its wealth management business (Bautzer and Nguyen, 2023). In the context of investments, Fieberg et al. (2023) argue that "publicly available information is the key driver of predictable asset price movements, which implies that AI applications such as GPT-4 may be particularly well-equipped to adjust portfolio allocation advice to the arrival of new information." OpenAI's efforts to integrate a browsing function may significantly change ChatGPT's abilities in this field, as current information from the web bypasses ChatGPT's knowledge cutoff in September 2021.[1]

Considering the vast array of information that needs to be evaluated and the limited attention of investors (Hirshleifer and Teoh, 2003), leveraging LLMs to assist in the information processing could be a game-changer in investment management. However, ChatGPT's responses are generated based on patterns observed in its training data, making it vulnerable to producing misinformation if presented with ambiguous or misleading prompts (Lipton and Steinhardt, 2018). It does not "understand" information in the human sense, rather it regurgitates patterns it has learned. This can sometimes lead to outputs that are plausible-sounding but factually incorrect.[2] Additionally, while the model can generate text on a wide range of topics, its depth of expertise on specific, nuanced subjects may vary, often reflecting the distribution and quality of its training data. To summarize, the potential strengths of the model to evaluate financial information comes from its ability to observe patterns in past data (the training data) and to apply these patterns to current information. If the same or very similar patterns lead to different outcomes, then the evaluation leads to misleading recommendations.

3. Related literature

Research on ChatGPT in financial contexts is still in its infancy. A few studies investigate whether ChatGPT can potentially assist portfolio construction. Ko and Lee (2023) focus on asset classes and find that ChatGPT's selections are better diversified and outperform random portfolios. Lopez-Lira and Tang (2023) use ChatGPT to assess the sentiment of single stock news headlines and find that portfolios based on ChatGPT sentiment scores outperform portfolios based on other LLMs such as BERT. Smales (2023) studies whether ChatGPT is able to classify monetary policy decisions made by the Reserve Bank of Australia.

Niszczota and Abbas (2023) investigate whether GPT can serve as a financial robo-advisor. The authors show that ChatGPT based on GPT-4 achieves a near-perfect 99% score in a financial literacy test which is a significant improvement over previous versions. Fieberg et al. (2023) design hypothetical investor profiles varying in their risk tolerance, age, and investment horizon. For those investors, the authors then ask ChatGPT to recommend a portfolio composition. They conclude that GPT's suggestions reflect an investor's individual circumstances such as their risk tolerance, risk capacity, and sustainability preferences but are insensitive to the investment horizon. Nevertheless, they conclude that the historical risk-adjusted performance is on par with a professionally managed benchmark portfolio.

Unlike the existing literature, we run a live experiment to examine a version of ChatGPT that autonomously browses the internet and weighs the relevance of the obtained information for its responses. We explore ChatGPT's performance in processing time-sensitive information and ranking stocks, which is a common practice of investors to navigate the vast array of investment options.

[1] At the time of our experiment, the OpenAI's browsing feature is still in beta and not available for ChatGPT users.

[2] Compared to previous versions, GPT-4 reduces the probability of hallucination, i.e., that the language model generates random answers not in line with reality.

- *Earnings surprise prompt:* Please evaluate all recent information up to today—note that this information includes relevant news outlets such as the Wall Street Journal but also investment-specific chat groups such as Reddit — on COMPANY NAME (TICKER SYMBOL). Based on this information, please provide a rating for the next earnings report on a scale from -5 to +5. Do you expect this earnings report to under- or overperform the current analyst forecasts? -5 indicates significant underperformance, 0 indicates an expected report in line with expectations, and +5 indicates significant outperformance.

- *Attractiveness prompt:* Based on the available information as of today, how do you assess the relative attractiveness of COMPANY NAME (TICKER SYMBOL) for investors intending to hold this stock in the following month? Compare the stock to all other S&P 500 stocks and assign a score on a scale from -5 to +5, where -5 indicates very unattractive, 0 indicates average attractiveness, and +5 indicates very attractive.

Fig. 1. ChatGPT prompts.
The figure summarizes the main prompt to collect earnings surprise and attractiveness ratings from ChatGPT.

Table 1
Summary statistics of ChatGPT's stock ratings.

	Mean	Std	Min	25%	50%	75%	Max
ChatGPT attractiveness rating	1.33	1.46	−5.0	0.5	1.0	2.0	5.0
ChatGPT earnings surprise rating	1.13	1.60	−5.0	0.0	1.0	2.0	5.0

The table reports summary statistics for ChatGPT earnings announcement ratings and ChatGPT attractiveness ratings.

4. Methodology

We study whether ChatGPT can support investors in making informed investment decisions. It is impossible to answer this question using historical data as it is not possible to make sure that ChatGPT does not use "future" data to make recommendations in backward looking analyses (look-ahead bias). Therefore, we conduct a live experiment. The setting of our tests is the 2023-Q2 earnings season. Between July 5, 2023 and September 8, 2023, we run two separate queries and ask ChatGPT to forecast the earnings announcement and evaluate the relative attractiveness of each S&P 500 firm. We summarize our prompts in Fig. 1.

Out-of-the-box, ChatGPT's knowledge is limited to information available up until September 2021 and it cannot browse the internet. Thus, it cannot provide responses to prompts that—like ours—require knowledge beyond that date. We insert additional information from a web search engine into our prompts to overcome this limitation. While such a browsing functionality is not yet supported by ChatGPT, OpenAI is already working on a similar approach. Thus, the ChatGPT responses we obtain are likely to resemble those obtained from future versions of ChatGPT.

We write an application that automatizes the Chrome web browser and loads the Web-ChatGPT extension. WebChatGPT is a third-party tool that interfaces with search engines to obtain relevant search results, combines prompts with the information obtained from the web, and asks ChatGPT for a response based on the newly acquired information. On a given date, our program iterates through the S&P 500 stocks and starts a conversation with ChatGPT by entering either our earnings surprise prompt or our attractiveness prompt. At this point, the WebChatGPT extension halts ChatGPT, opens a separate browser session in the background, and asks ChatGPT to formulate a Yahoo search query which is likely to retrieve relevant information for our prompt. Based on this query, the program obtains the URLs of the top ten Yahoo search results within the last four weeks. For each URL, the application crawls its contents, opens separate browser sessions in the background, and asks ChatGPT to generate content summaries of the websites. The summaries are then fed into a modified version of our original prompt as summarized in Fig. 2.

Our analysis is based on ChatGPT's responses to the modified prompt. We use Refinitiv Eikon to obtain the list of S&P 500 constituents and market-related, fundamental, and IBES analyst data. We obtain daily news volume and news sentiment scores from Bloomberg.

5. Results

We first describe the ratings that ChatGPT issues for earnings announcements and attractiveness. Table 1 shows that the average earnings announcement and attractiveness ratings are both positive, with a mean (median) of 1.13 (1) and 1.33 (1), respectively. Over our sample period, ChatGPT makes use of the complete range of possible ratings. However, the range of ratings varies over time. Fig. 3 shows that ChatGPT uses the full range of possible ratings only in the second half of our sample period. Initially, especially (very) low ratings are not used.

Forget all previous instructions. Pretend to be an investment professional. I will give you a question or an instruction. Your objective is to answer my question or fulfill my instruction.
My question or instruction is: [placeholder for earnings surprise prompt or attractiveness prompt]
For your reference, today's date is [placeholder for date and time].
It's possible that the question or instruction, or just a portion of it, requires relevant information from the internet to give a satisfactory answer or complete the task. I'm providing you with the necessary information already obtained from the internet below. This sets the context for addressing the question or fulfilling the instruction, so you don't need to access the internet to answer my question or fulfill my instruction. Write a reply to the given question or instruction using the information provided below in the best way you can.
A strict requirement for you is to only provide the score. Do not use complete sentences and do not provide additional explanations or other content.
Now, reply to the given question or instruction with this information:
"""

NUMBER:[Placeholder for the ID of the Yahoo search result]
URL: [Placeholder for the URL]
TITLE: [Placeholder for the title of the website]
CONTENT: [Placeholder for the summarized content of the website]

...
"""

Fig. 2. Modified prompt.
The figure summarizes the modified prompt to collect earnings surprise and attractiveness ratings, combined with information that WebChatGPT obtains from Yahoo web search results.

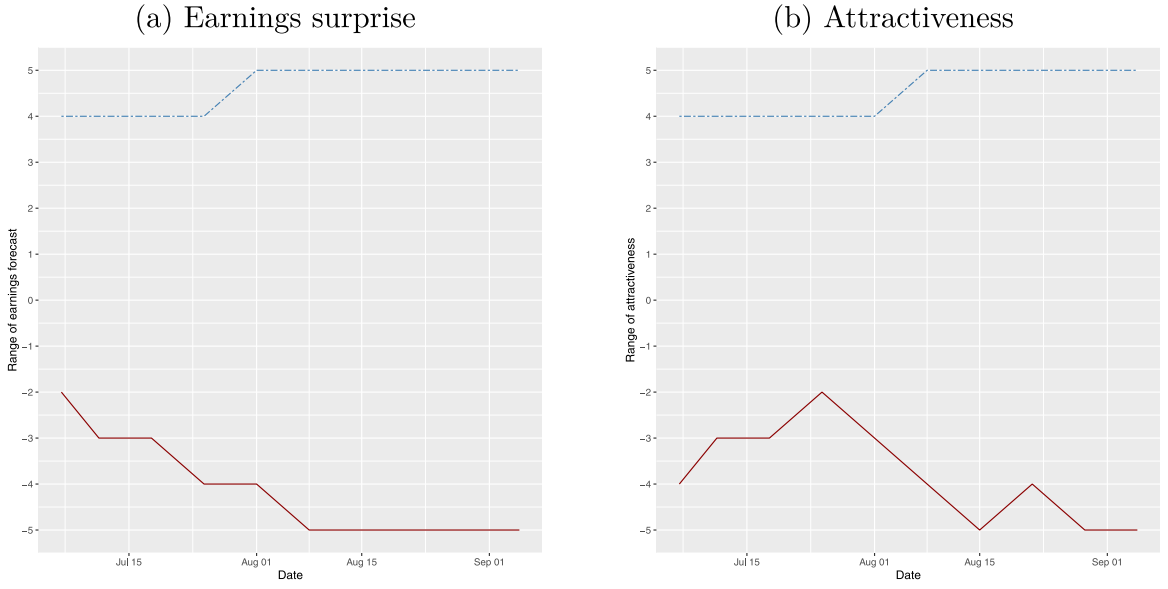
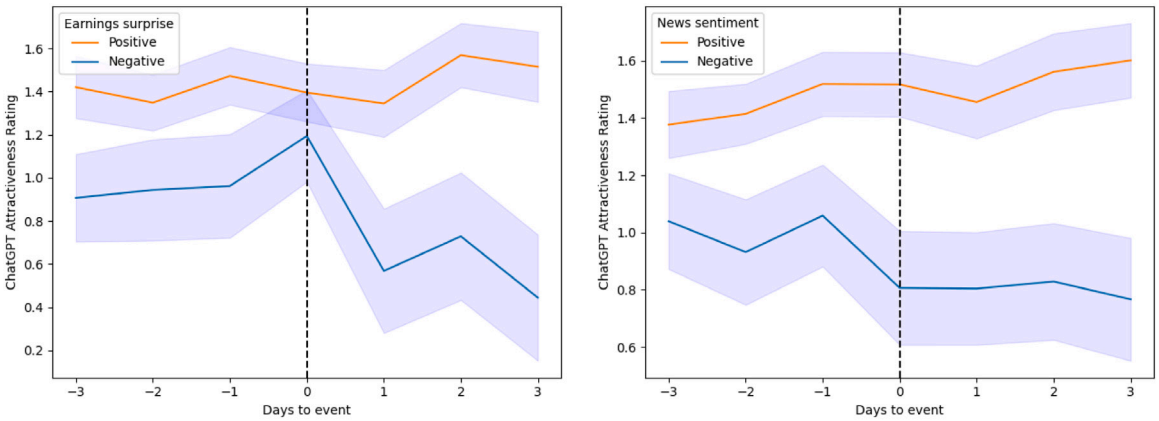


Fig. 3. Min–max plot of ChatGPT ratings.
The figure shows the minimum and maximum ChatGPT rating scores about a stock's earnings surprise forecast (Panel a) and a stock's relative attractiveness (Panel b) for each week between 2023-07-05 and 2023-09-08.



(a) Earnings announcements (b) News events

Fig. 4. ChatGPT rating changes in response to earnings announcements and news events.
The figure shows the response of ChatGPT attractiveness ratings to positive and negative earnings surprises (Panel a) and to positive and negative news events (Panel b). ChatGPT attractiveness ratings range from −5 to +5 and evaluate the relative desirability of investing in a particular company's stock over the next month in comparison to all other S&P 500 stocks, with −5 indicating a very unattractive investment, 0 indicating an average level of attractiveness, and +5 indicating a very attractive investment opportunity. Earnings surprises are defined as the average deviation between analyst EPS consensus forecasts and actual EPS relative to the stock price. News events are defined as days where the number of news articles is three standard deviations above the mean number of news articles over the past 30 days. Bloomberg news sentiment scores determine positive or negative news events.

Table 2
GPT-4 earnings forecasts and earnings announcements.

	(1)	(2)	(3)	(4)	(5)
Sample	Full	Full	Full	Low SD	High SD
Dep. var.	EPS Actual				
(Intercept)	0.0152	−0.0007	−0.0008	0.0009	−0.0019
	(21.1269)	(−2.1106)	(−2.6442)	(3.1633)	(−3.1002)
Rating GPT	0.0000	0.0003	0.0003	−0.0002	0.0007
	(0.1056)	(2.5873)	(2.5773)	(−1.8410)	(3.3482)
EPS forecast (mean)		1.0706		0.9997	1.1144
		(64.1391)		(63.8086)	(41.4585)
EPS forecast (median)			1.0866		
			(68.3991)		
Obs.	459	459	459	235	224
Adj. R^2	−0.0022	0.8998	0.9108	0.9456	0.8851

The table presents the results of regressions on actual earnings (EPS Actual). We regress the last GPT-4 earnings announcement rating before each announcement (Rating GPT) on actual earnings. EPS forecast denotes the mean (median) analyst forecast for the announcement. Columns (1)-(3) report regressions on the full sample; Column (4) includes stocks with below median standard deviation of analyst forecasts; Column (5) includes stocks with above median standard deviation of analyst forecasts. Analyst forecast data is from IBES. t-statistics are reported in parentheses.

Table 2 summarizes correlations between GPT-4 earnings forecasts and actual earnings using cross-sectional OLS regressions. Column 1 indicates that we do not find any correlation between earnings forecasts and earnings. However, after controlling for analyst forecasts, we find a positive correlation between GPT ratings and actual earnings (Columns 2 and 3, p-values < 0.01). Note that we explicitly ask ChatGPT to provide earnings forecasts beyond current analyst forecasts ("Do you expect this earnings report to under- or overperform the current analyst forecasts?"). Our results indicate that ChatGPT is able to execute that task.

In columns 4 and 5, we separately investigate earnings announcements with low and high dispersion in analyst forecasts, as measured by the standard deviation (SD) of analyst forecasts. We find that the rating of ChatGPT significantly correlates with actual earnings in particular for high dispersion analyst forecasts ($p \approx 0.01$).

Next, we investigate whether ChatGPT can provide valuable assistance in evaluating the vast amount of information on particular stocks. Fig. 4 shows the reaction of ChatGPT attractiveness ratings to earnings announcements in particular and to news in general. In Panel a, we focus on earnings announcements. We plot the average attractiveness rating for positive and negative earnings surprises. We define earnings surprises as the average deviation between analyst EPS consensus forecasts and actual EPS relative to the stock price. Noteworthy is the clear distinction between positive and negative earnings announcements. Stocks that exhibit positive earnings surprises receive significantly higher attractiveness ratings compared to stocks that realize negative earnings surprises.

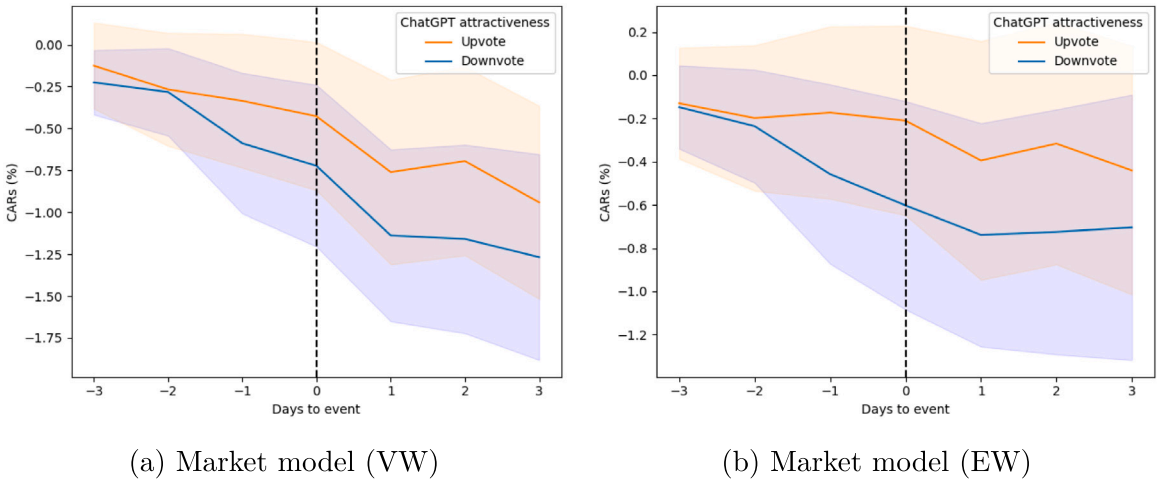


Fig. 5. Event study on ChatGPT attractiveness rating changes.
The figure shows the cumulative abnormal returns (CARs) around significant changes in ChatGPT attractiveness ratings on date t. A significant rating change is determined by a change in ratings of at least 4 scores. CARs are computed within event window $[t-7, t+7]$ and are based on a market model estimated in $[t-259, t-10]$. In Panel a, abnormal returns are based on value-weighted S&P 500 index returns (VW). In Panel b, abnormal returns are based on equal-weighted S&P 500 returns (EW). Shaded areas indicate 95%-confidence intervals using standard errors averaged across events and firms.

Table 3
Portfolio sorts using ChatGPT attractiveness ratings.

	Least attractive	2	3	4	Most attractive	5-1
Mean return	−0.15	−0.08	−0.11	−0.07	−0.08	0.07
	(−13.74)	(−7.87)	(−10.82)	(−6.18)	(−6.57)	(4.35)
Sharpe ratio	−0.26	−0.13	−0.16	−0.10	−0.13	0.17

The table shows daily average returns (in %) for the next 30 days following ChatGPT's attractiveness rating together with the Sharpe ratios of rating quintile portfolios. Quintile portfolios are formed at the end of each week within our sample period and are based on the weekly average attractiveness rating provided by ChatGPT. t-statistics are reported in parentheses.

Moreover, ChatGPT significantly adjusts its attractiveness ratings following negative earnings surprises by approximately half a point. We do not observe an adjustment to positive earnings surprises.

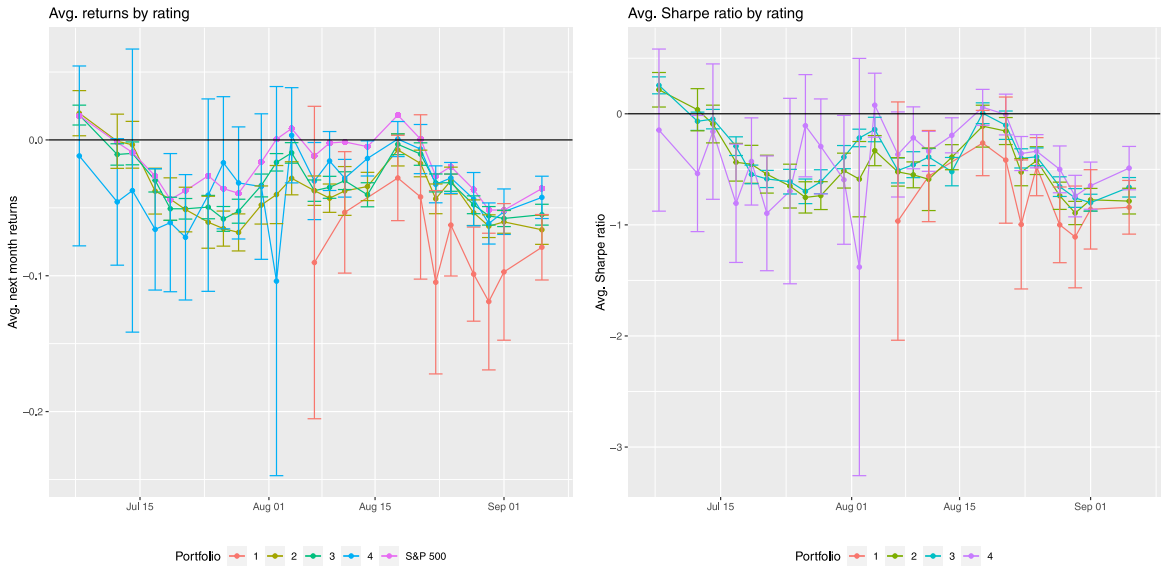
Panel b of Fig. 4 studies the more general case of news events. Again, we observe the clear distinction between positive and negative news events, as measured by Bloomberg news sentiment scores. Interestingly, ChatGPT adjusts ratings for negative news events immediately before the news event, which may be explained by news leakages or rumors. Overall, these results provide evidence in support of the notion that ChatGPT is able to evaluate news, at least to some degree.

Next, we study stock returns around large rating changes. Fig. 5 summarizes the results of an event study on changes in ChatGPT's attractiveness ratings of at least 4 points. While we do not observe statistically significant differences between positive changes in the attractiveness ratings (upvotes) and negative changes in the attractiveness ratings (downvotes), on average, the spread between abnormal stock returns of upvotes and downvotes widens and the standard deviation of abnormal stock returns increases. As the spread begins to open at $t = -2$, it seems that ChatGPT adjusts its ratings in response to the stock market.

Finally, we turn to the relationship between GPT's attractiveness ratings and future stock returns and Sharpe ratios. Every time we collect attractiveness ratings, we form portfolios based on those ratings. Then, we calculate the returns (Sharpe ratios) of those equally-weighted portfolios over the next 30 days. We plot the average next-month returns of those portfolios in Fig. 6. While the portfolio with the highest ratings provides lower returns at the beginning of our sample period, higher rated portfolios yield higher returns in particular towards the end of our sample period (see also Fig. 7) and over the full sample period, on average (see Table 3). Fig. 7 visualizes the results of simple cross-sectional OLS regressions that we run every time we collect attractiveness ratings. We regress stock returns (Sharpe ratios) over the next 30 days on the attractiveness ratings and plot the estimated coefficients together with 95%-confidence intervals.

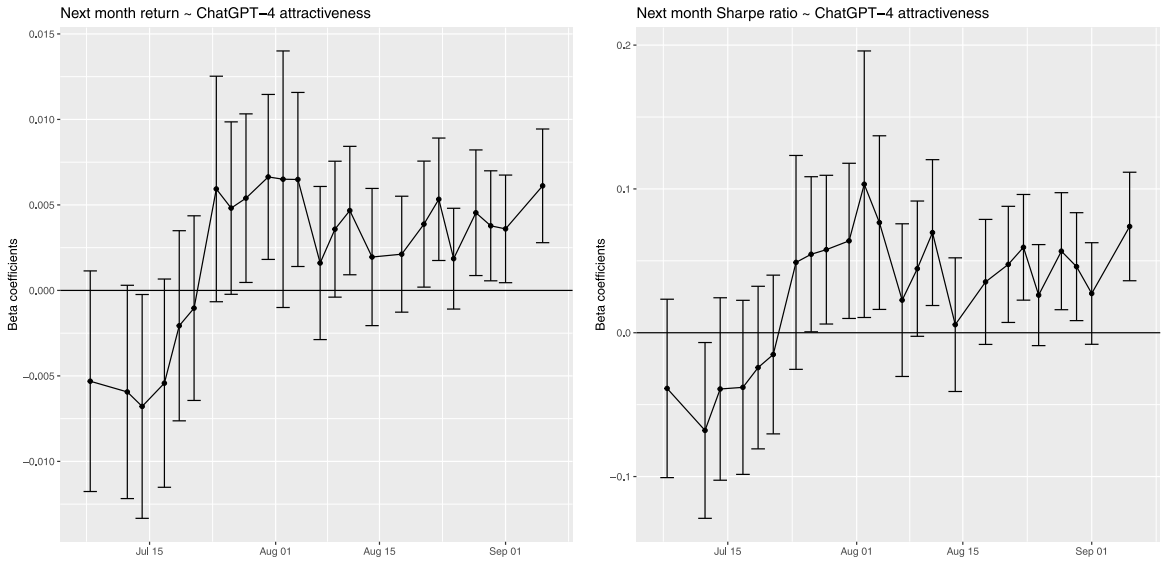
6. Discussion

LLMs have aroused enormous interest lately. The rapid integration of LLMs and similar technologies in various industries highlight the hopes and expectations placed in these technologies. In this paper, we consider the use of such technologies in the field of investment. We study whether ChatGPT can provide valuable assistance in the processing of information and the selection of stocks.



(a) Return forecasts (b) Sharpe ratio forecasts

Fig. 6. Avg. next-month returns of portfolios based on ChatGPT-4 ratings.
The figure shows the average next-month returns (Panel A, together with the next-month return of an investment into the S&P) and Sharpe ratios (Panel B) of 4 portfolios formed according to ChatGPT-4 attractiveness ratings. Every time ChatGPT attractiveness ratings are collected, portfolios are formed based on the attractiveness scores. Portfolio 1 includes stocks with an attractiveness rating ≤ -3, portfolio 2 includes ratings > -3 and ≤ 0, portfolio 3 includes ratings > 0 and ≤ 3, and portfolio 4 includes ratings > 3. Portfolios have to include at least 5 stocks. We then plot the equal-weighted average of portfolio returns (Sharpe ratios) over the next 30 days. Dates show the time of portfolio formation.



(a) Return forecasts (b) Sharpe ratio forecasts

Fig. 7. Correlations between ChatGPT attractiveness ratings and future stock returns.
The figure plots coefficients of an OLS regression of next-month returns (Panel A) and next-month Sharpe ratios (Panel B) on ChatGPT-4 attractiveness ratings together with 95%-confidence intervals. Every time ChatGPT attractiveness ratings are collected, we run one cross-sectional regression of stock returns (Sharpe ratios) over the next 30 days on the attractiveness ratings. Dates show the time of ratings collection.

In a live experiment, we find that ratings of stocks by ChatGPT positively correlate to future (out-of-sample) stock returns. If the market is at least weakly efficient and ChatGPT would assign ratings to stocks randomly, all portfolios based on those ratings should provide approximately the same return, a zero abnormal return, on average. However, our results indicate that the assignment of ratings to stocks is not random. Instead, ChatGPT seems to be able to successfully identify stocks that yield superior performance over the next month.

ChatGPT-4 seems to have some ability to evaluate news information and summarize its evaluation into a simple score. We find clear evidence that ChatGPT is able to distinguish between positive and negative news events, and adjusts its recommendation following negative news.

Notwithstanding our results, we advice caution in using such technologies for investment decisions without thoroughly reviewing the suggestions. LLMs like ChatGPT do not deliver complete factual accuracy. Factual accuracy increases over time; however, both our results and the previous literature (Chen et al., 2023) show that the performance of, for example, ChatGPT varies over time. Thus, LLMs, like ChatGPT, might not always provide accurate or reliable financial advice.

CRediT authorship contribution statement

Matthias Pelster: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Joel Val:** Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

References

Ali, Ashiq, Chen, Xuanjuan, Yao, Tong, Yu, Tong, 2020. Can mutual funds profit from post earnings announcement drift? The role of competition. J. Bank. Financ. 114, 105774.

Allee, Kristian D., Do, Chuong, Sterin, Mikhail, 2021. Product market competition, disclosure framing, and casting in earnings conference calls. J. Account. Econ. 72 (1), 101405.

Barber, Brad M., Odean, Terrance, 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. Rev. Financ. Stud. 21 (2), 785–818.

Bautzer, Tatiana, Nguyen, Lananh, 2023. Morgan Stanley to launch AI chatbot to woo wealthy. Thomson Reuters, https://www.reuters.com/technology/morgan-stanley-launch-ai-chatbot-woo-wealthy-2023-09-07. (Accessed 12 November 2023).

Brown, Tom B., Mann, Benjamin, Ryder, Nick, Subbiah, Melvin, Kaplan, Jared, Dhariwal, Prafulla, Neelakantan, Arvind, Shyam, Pranav, Sastry, Girish, Askell, Amanda, et al., 2020. Language models are few-shot learners. In: Advances in Neural Information Processing Systems, Vol. 33.

Chen, Lingjiao, Zaharia, Matei, Zou, James, 2023. How is ChatGPT's behavior changing over time? arXiv preprint arXiv:2307.09009.

Cucchiara, Rita, 2023. What large language models like GPT can do for finance. Nature Italy.

Dowling, Michael, Lucey, Brian, 2023. ChatGPT for (Finance) research: The Bananarama conjecture. Finance Res. Lett. 53, 103662.

Doyle, Jeffrey T., Lundholm, Russell J., Soliman, Mark T., 2006. The extreme future stock returns following I/B/E/S earnings surprises. J. Account. Res. 44 (5), 849–887.

Fieberg, Christian, Hornuf, Lars, Streich, David J., 2023. Using GPT-4 for financial advice. In: SSRN Working Paper.

Hirshleifer, David, Teoh, Siew Hong, 2003. Limited attention, information disclosure, and financial reporting. J. Account. Econ. 36 (1), 337–386.

Kahneman, Daniel, 1973. Attention and Effort. Prentice Hall.

Ko, Hyungjin, Lee, Jaewook, 2023. Can ChatGPT improve investment decision? From a portfolio management perspective. In: SSRN Working Paper.

Lipton, Zachary C., Steinhardt, Jacob, 2018. Troubling trends in machine learning scholarship. arXiv preprint arXiv:1807.03341.

Lopez-Lira, Alejandro, Tang, Yuehua, 2023. Can ChatGPT forecast stock price movements? Return predictability and large language models. In: SSRN Working Paper.

Milian, Jonathan A., 2015. Unsophisticated arbitrageurs and market efficiency: Overreacting to a history of underreaction? J. Account. Res. 53 (1), 175–220.

Niszczota, Pawel, Abbas, Sami, 2023. GPT has become financially literate: Insights from financial literacy tests of GPT and a preliminary test of how people use it as a source of advice. Finance Res. Lett. 58, 104333.

Noh, Joonki, Zhou, Dexin, 2022. Executives' Blaming external factors and market reactions: Evidence from earnings conference calls. J. Bank. Financ. 134, 106358.

Rennekamp, Kristina M., Sethuraman, Mani, Steenhoven, Blake A., 2022. Engagement in earnings conference calls. J. Account. Econ. 74 (1), 101498.

Smales, Lee A., 2023. Classification of RBA monetary policy announcements using ChatGPT. Finance Res. Lett. 58, 104514. http://dx.doi.org/10.1016/j.frl.2023.104514.

Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N., Kaiser, Łukasz, Polosukhin, Illia, 2017. Attention is all you need. In: Guyon, I., Luxburg, U. Von, Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Vol. 30. Curran Associates, Inc..