Can LLMs be Good Financial Advisors?: An Initial Study in Personal Decision Making for Optimized Outcomes

**Kausik Lakkaraju, Sai Krishna Revanth Vuruma, Vishal Pallagani,
Bharath Muppasani, Biplav Srivastava**

University of South Carolina

Abstract

Increasingly powerful Large Language Model (LLM) based chatbots, like ChatGPT and Bard, are becoming available to users that have the potential to revolutionize the quality of decision-making achieved by the public. In this context, we set out to investigate how such systems perform in the personal finance domain, where financial inclusion has been an overarching stated aim of banks for decades. We asked 13 questions representing banking products in personal finance: bank account, credit card and certificate of deposits and their inter-product interactions, and decisions related to high-value purchases, payment of bank dues, and investment advice, and in different dialects and languages (English, African American Vernacular English, and Telugu). We find that although the outputs of the chatbots are fluent and plausible, there are still critical gaps in providing accurate and reliable financial information using LLM-based chatbots.

Introduction

Consider a freshman that has just started making personal financial decisions. They open a bank account to save up money and get their first credit card. They are given some seed money by their family and they also start earning by working on campus. The student is encouraged by their support system to start thinking about saving into products like Certificate of Deposits (CDs) that earn higher interest. As the student makes a series of decisions in their academic and subsequent professional life, they need to make sound financial decisions and may look for resources online to assist them. An optimal decision needs to consider how the banking products interact with each other along with the changing needs of the student.

For users like this student, increasingly powerful LLM-based chatbots that have the potential to revolutionize the quality of decision for personal finance are becoming available. LLMs have demonstrated tremendous potential across diverse domains (Zhao et al. 2023), such as natural language processing (Min et al. 2021) and protein structure (Hu et al. 2022), and have been claimed to show sparks of artificial general intelligence (Bubeck et al. 2023). These models have been implemented in several applications, ranging from mental health assistants (Xiao et al. 2023) to financial advisement (Yue and Au 2023). In the finance domain, LLMs have been used to develop applications such as fraud detection, risk management, and financial forecasting (Alberto Pozanco 2022). They have been used to analyze financial data, predict stock prices, and generate automated reports. However, with the advent of recent models such as OpenAI's ChatGPT, Google's Bard, and BloombergGPT (Wu et al. 2023), a comparative chatbot study is needed to evaluate their ability to be financial advisors. In this paper, we present an initial study of ChatGPT and Bard in providing personal decision-making for optimized outcomes.

It is widely known that LLMs based systems have unique limitations. For example, they may struggle with common-sense reasoning tasks (Li et al. 2022), encounter challenges when handling symbols (Frieder et al. 2023), and are susceptible to hallucinations (Bang et al. 2023).

With this work, we make the following contributions:

- identify a personal financial planning scenario involving a series of tasks (plans) and optimization of decisions.
- show how leading LLM-based chatbots perform in them and analyze their behavior.
- lay out challenges that future chatbots in this area should overcome to provide trusted financial recommendations.

We thus highlight the potential and limitations of current LLM-based systems - ChatGPT and Bard - in their role as financial advisors. We included all the queries posed and responses from both ChatGPT and Bard in our GitHub repository[1] along with a few snapshots of the actual conversations.

---

[1] https://github.com/ai4society/LLM-CaseStudies/tree/main/Finance

Product Interactions	Query Identifier	Queries	Variables with their values	Constraints
CC	Q1	I am making a **purchase of $1000** using my credit card. My **billing cycle is from March 25th to April 24th**. Today is March 31st, and I have a **due of $2000** on my account. My total **credit line is $2,800**. Would you recommend I make the purchase now or later in the future?	$x_{PA} = 1000$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2000$, $x_{CL} = 2800$	$x_{DA} + x_{PA} < x_{CL}$ (1)
	Q2	I am making a **purchase of $1000** using my credit card. My **billing cycle is from March 25th to April 24th**. Today is March 31st, and I have a **due of $2000** on my account. My total **credit line is $3,800**. Would you recommend I make the purchase now or later in the future?	$x_{PA} = 1000$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2000$, $x_{CL} = 3800$	
	Q3	I get **5% cashback** if I buy furniture using my credit card. I am **buying furniture worth $1000** using my credit card. My **billing cycle is from March 25th to April 24th**. Today is March 31st, and I have a **due of $2000** on my account. My total **credit line is $2,800**. Would you recommend I make the purchase now or later in the future?	$x_{CP} = 5\%$, $x_{PA} = 1000$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2000$, $x_{CL} = 2800$	
	Q4	I get **5% cashback** if I buy furniture using my credit card. I am **buying furniture worth $1000** using my credit card. My **billing cycle is from March 25th to April 24th**. Today is March 31st, and I have a **due of $2000** on my account. My total **credit line is $3,800**. Would you recommend I make the purchase now or later in the future?	$x_{CP} = 5\%$, $x_{PA} = 1000$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2000$, $x_{CL} = 3800$	
CC (AAVE)	Q5	I be makin' a **purchase of $1000** usin' i's credit card. I's **billin' cycle be from march 25th to april 24th**. Today be march 31ts, and i done a **due of $2000** on i's account. I's total **credit line be $2,800**. Would you recommend i make de purchase now o lateh in de future?	$x_{PA} = 1000$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2000$, $x_{CL} = 2800$	
CC (Telugu)	Q6	నేను నా క్రెడిట్ కార్డ్ని ఉపయోగించి **$1000 కొనుగోలు** చేస్తున్నాను. నా **బిల్లింగ్ సైకిల్ మార్చి 25 నుండి ఏప్రిల్ 24 వరకు** ఉంది. ఈరోజ మార్చి 31, మరియు నా ఖాతాలో **$2000 బకాయి** ఉంది. నా మొత్తం **క్రెడిట్ లైన్ $2,800**. నేను ఇప్పుడు లేదా భవిష్యత్తులో కొనుగోలు చేయాలని మీరు సిఫార్సు చేస్తారా?	$x_{PA} = 1000$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2000$, $x_{CL} = 2800$	
	Q7	నేను నా క్రెడిట్ కార్డ్ని ఉపయోగించి **$1000 కొనుగోలు** చేస్తున్నాను. నా **బిల్లింగ్ సైకిల్ మార్చి 25 నుండి ఏప్రిల్ 24 వరకు** ఉంది. ఈరోజ మార్చి 31, మరియు నా ఖాతాలో **$2000 బకాయి** ఉంది. నా మొత్తం **క్రెడిట్ లైన్ $3,800**. నేను ఇప్పుడు లేదా భవిష్యత్తులో కొనుగోలు చేయాలని మీరు సిఫార్సు చేస్తారా?	$x_{PA} = 1000$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2000$, $x_{CL} = 3800$	
CC and AB	Q8	I am making a **purchase of $1000** using my credit card. My **billing cycle is from March 25th to April 24th**. Today is March 31st, and I have a **due of $2000** on my account. My total **credit line is $3,800**. I have **$10,000 in my bank** which I can use to pay my credit card balance any time. Would you recommend I make the purchase now or later in the future?	$x_{PA} = 1000$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2000$, $x_{CL} = 3800$, $x_{AB} = 10000$	Constraint (1) must be satisfied. In addition, if the user chooses to pay the due immediately, the following constraints must also hold true. $x_{DA} < x_{AB}$ (2) , $x_{PA} < x_{CL}$ (3)
	Q9	I get **5% cashback** if I buy furniture using my credit card. I am **buying furniture worth $1000** using my credit card. My **billing cycle is from March 25th to April 24th**. Today is March 31st, and I have a **due of $2000** on my account. My total **credit line is $3,800**. I have **$10,000 in my bank** which I can use to pay my credit card balance any time. Would you recommend I make the purchase now or later in the future?	$x_{CP} = 5\%$, $x_{PA} = 1000$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2000$, $x_{CL} = 3800$, $x_{AB} = 10000$	
CC and CD	Q10	I have a credit card **due of $2800**. The total **credit line is $2800**. If I don't pay a **minimum of $100** by the end of billing cycle, my **APR would be 27%**. If I pay the minimum amount by the end of billing cycle, **APR will be 25%**. My **billing cycle is from March 25th to April 24th**. Today is March 31st. If I choose to deposit some amount as certificate of deposit (CD), I will get an **interest of 6% on the amount deposited**. Do you recommend I pay the full credit card due or do a certificate of deposit or pay my due and deposit the rest?	$x_{APR} = 27\%$ (with late fee) and 25% without late fee, $x_{MD} = 100$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2800$, $x_{CL} = 2800$, $x_{CDP} = 6\%$	$x_{DA} < x_{CL}$ (4) AB was not provided in this query. So we cannot specify any additional constraints in this case from the given data.
	Q11	I have a credit card **due of $2800**. The total **credit line is $3800**. If I don't pay a **minimum of $100** by the end of billing cycle, my **APR would be 27%**. If I pay the minimum amount by the end of billing cycle, **APR will be 25%**. My **billing cycle is from March 25th to April 24th**. Today is March 31st. If I choose to deposit some amount as certificate of deposit (CD), I will get an **interest of 6% on the amount deposited**. Do you recommend I pay the full credit card due or do a certificate of deposit or pay my due and deposit the rest?	$x_{APR} = 27\%$ (with late fee) and 25% without late fee, $x_{MD} = 100$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2800$, $x_{CL} = 3800$, $x_{CDP} = 6\%$	
CC, CD and AB	Q12	I have a credit card **due of $2800**. The total **credit line is $2800**. If I don't pay a **minimum of $100** by the end of billing cycle, my **APR would be 27%**. If I pay the minimum amount by the end of billing cycle, **APR will be 25%**. My **billing cycle is from March 25th to April 24th**. Today is March 31st. I currently have **$2,800 in my personal checking account**. If I choose to deposit some amount as certificate of deposit (CD), I will get an interest of **6% on the amount deposited**. Do you recommend I pay the full credit card due with my personal account balance or do a certificate of deposit or pay my due and deposit the rest?	$x_{APR} = 27\%$ (with late fee) and 25% without late fee, $x_{MD} = 100$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2800$, $x_{CL} = 2800$, $x_{CDP} = 6\%$, $x_{AB} = 2800$	$[(x_{DA} - x_{MD}) * x_{APR} \leq (x_{AB} - x_{MD}) * x_{CDP}]$ (5) , $[(x_{AB} - x_{DA}) > 0]$ (6)
	Q13	I have a credit card **due of $2800**. The total **credit line is $2800**. If I don't pay a **minimum of $100** by the end of billing cycle, my **APR would be 27%**. If I pay the minimum amount by the end of billing cycle, **APR will be 25%**. My **billing cycle is from March 25th to April 24th**. Today is March 31st. I currently have **$3,800 in my personal checking account**. If I choose to deposit some amount as certificate of deposit (CD), I will get an **interest of 6% on the amount deposited**. Do you recommend I pay the full credit card due with my personal account balance or do a certificate deposit or pay my due and deposit the rest?	$x_{APR} = 27\%$ (with late fee) and 25% without late fee, $x_{MD} = 100$, x_{BC} = (March 25th - April 24th), $x_{DA} = 2800$, $x_{CL} = 2800$, $x_{CDP} = 6\%$, $x_{AB} = 3800$	

Table 1: Table showing different product interaction categories considered, query identifiers, the queries posed to the chatbots under each category, variables used in each query with their corresponding chosen values and constraints the chatbots need to consider while answering the user queries.

Personal Finance Use Case

Setup: Tools and Procedure

Chatbots Tested

1. **ChatGPT:** ChatGPT (OpenAI 2023) is an LLM-based chatbot created by OpenAI that was trained on large amount of text data from the internet, including books and articles. ChatGPT is capable of answering questions, generating text and converse with users in a natural way. It can also learn from users and adapt to new information.

2. **Bard:** Bard (Google 2023) is an LLM-based chatbot created by Google that was trained on large amount of text data and is capable of generating human-like text in response to user prompts and queries. Like ChatGPT, it is also capable of conversing with users about wide variety of topics in a natural way and adapt to new information.

Product Interaction Categories Product interaction refers to interaction between different products like Credit Card (CC), Certificate of Deposit (CD) and Account Balance (AB). Each product has different quantitative properties. For example, credit card due, limit line and billing cycle are some of the properties that would provide credit card information (not private information) of the user. Different properties pertaining to these products are:

- **Purchase Amount (PA)**: It is the amount spent by the user on purchase of a product.
- **Billing Cycle (BC)**: It is the billing cycle of user's credit card.
- **Due Amount (DA)**: The amount that is due on the user's credit card for the specified billing cycle.
- **Credit Line (CL)**: The maximum amount that user could spend using their credit card. If the amount spent exceeds this value, the credit card company could charge additional interest.
- **Cashback Percentage (CP)**: The % of amount which will be returned to the user in the form of cashback on buying furniture using their credit card.
- **Account Balance (AB)**: The amount of cash present in user's personal bank account.
- **Annual Percentage Rate (APR)**: The APR is charged if there is due on the credit card after the due date. Some financial institutions choose to charge a late fee if the **minimum due (MD)** is not paid. It is calculated by the formula, Daily Period Rate (DPR) x Billing Cycle (in days) x Average Daily Balance (ADB).
- **Certificate of Deposit Percentage (CDP)**: The % of interest accumulated on the cash deposited by the user in the form of CD.

Based on different combinations of these products, we classified the queries into 4 categories. These four categories along with the queries posed under each category, the variables used in each query and the constraints the chatbot has to take into consideration to make a sound recommendation are shown in Table 1. In the CC category, we considered a different dialect of English called African American Vernacular English (AAVE) and Telugu, one of the well-known

S.No.	Bard	ChatGPT
1.	Bard gives accurate results if the question is asked directly (for ex., $2,250 x 0.0006849 x 30 = $46.23075.)	ChatGPT gives inaccurate results if the question is asked directly ($2,250 x 0.0006849 x 30 = $46.90 (rounded to the nearest cent))
2.	Bard does not utilize the information the user provides completely and calculates CUR less often than ChatGPT.	ChatGPT calculates CUR and reasons using the computed CUR more often than Bard
3.	Bard usually does not give personalized suggestions (especially, when the (Due + purchase amount) > Credit line).	ChatGPT gives personalized suggestions more often than Bard.
4.	As a response to one of the queries, Bard gave a recommendation by making use of a table with different options that user could choose from as shown in Figure 1.	ChatGPT did not use any kind of visual aids.
4.	Bard gave biased recommendation i.e., biased towards recommending the user to make the purchase immediately (in one case, it gave only pros for buying the furniture immediately even though it has serious cons).	ChatGPT never gave biased recommendations (it never encourages the user to buy the furniture immediately unless there is no risk involved).
5.	Bard gives 3 different drafts (with some changes in the response) for the same query.	ChatGPT does not provide different drafts.
6.	With each query posed, the content (calculations) of Bard is not improving as much as ChatGPT. It is not learning from its mistakes immediately.	ChatGPT corrects its errors more often than Bard
7.	Bard understood African-American Vernacular English (AAVE) dialect and gave a reasonable response to the query	When query was posed in AAVE dialect, ChatGPT did not understand it immediately. When we posed the same query again in the same dialect, it understood the query and gave a reasonable recommendation .
8.	Bard was not trained to understand Telugu language.	Though ChatGPT can understand Telugu language and responds in Telugu if the user query is in Telugu, the response it generated was incomplete and had a lot of grammatical errors which made the response very hard to understand.

Table 2: Differences between the responses generated by Bard and ChatGPT when queries related to finance domain were posed.

languages from India, to observe how the chatbots handle queries in a different language or dialect.

Findings

In this subsection, we present the findings from the interesting (and sometimes insightful) conversations we had with Bard and ChatGPT.

Differences Between the Chatbots Table 2 shows the differences that were identified between Bard and ChatGPT when queries listed out in Table 1 were asked. We compare these models on various criteria related to their performance in answering queries. The criteria include accuracy, utilization of user information, personalized suggestions, use of visual aids, bias in recommendations, provision of multiple response drafts, learning from mistakes, and understanding of different dialects and languages.

Error Categories We identified some limitations / errors in the responses generated by both the chatbots and classified them into the following categories:

- **Lack of Personalized Recommendations:** When the agent makes a generalized recommendation without using all the information provided by the user, we consider this as lack of personalized recommendation.

I have a credit card due of $2800. The total credit line is $3800. If I don't pay a minimum of $100 by the end of billing cycle, my APR would be 27%. If I pay the minimum amount by the end of billing cycle, APR will be 25%. My billing cycle is from March 25th to April 24th. Today is March 31st. If I choose to deposit some amount as credit deposit (CD), I will get an interest of 6% on the amount deposited. Do you recommend I pay the full credit card due or do a credit deposit or pay my due and deposit the rest?

Given the information you have provided, I recommend that you pay the full credit card due of $2800. Here are a few reasons why:

- If you don't pay the full amount, you will be charged interest at a rate of 27%. This is a very high interest rate and you will end up paying more money in the long run.
- If you pay the minimum amount, your APR will be 25%. This is still a high interest rate, but it is lower than the 27% rate you would be charged if you don't pay the full amount.
- If you deposit some amount as a credit deposit, you will only earn interest on the amount deposited. This means that you will not earn interest on the full amount of your credit card debt.
- If you pay your due and deposit the rest, you will still be charged interest on the amount you don't pay. This means that you will end up paying more money in the long run.

The best way to save money on interest is to pay your credit card balance in full each month. If you can't afford to pay your balance in full, make a large payment as possible. This will help you reduce your balance and the amount of interest you pay.

If you are struggling to manage your credit card debt, consider talking to a credit counselor. They can help you develop a plan to get out of debt.

Here is a table that shows how much interest you would pay on your credit card debt if you paid the minimum amount, the full amount, or deposited some amount as a credit deposit:

Payment Option	Interest Charged	Total Amount Paid
Pay minimum amount	$784.40	$3584.40
Pay full amount	$0	$2800
Deposit some amount as a credit deposit	$168	$3062

As you can see, you would save the most money by paying the full amount of your credit card debt.

Additionally, if you deposited some amount as a credit deposit, you would still be charged interest on the amount you owe on your credit card. This means that you would end up paying more money in the long run.

Figure 1: Bard Response (Q11-AB11).

I have a credit card due of $2800. The total credit line is $3800. If I don't pay a minimum of $100 by the end of billing cycle, my APR would be 27%. If I pay the minimum amount by the end of billing cycle, APR will be 25%. My billing cycle is from March 25th to April 24th. Today is March 31st. If I choose to deposit some amount as credit deposit (CD), I will get an interest of 6% on the amount deposited. Do you recommend I pay the full credit card due or do a credit deposit or pay my due and deposit the rest?

Based on the information you provided, I would recommend paying off the full credit card balance of $2800 before the end of the billing cycle to avoid paying any interest charges. Here's why:

1. High APR: If you don't pay the minimum payment of $100 by the end of the billing cycle, you will be charged a high APR of 27%. This can quickly add up and increase the amount you owe.
2. Lower APR with minimum payment: If you make the minimum payment of $100 by the end of the billing cycle, your APR will be 25%. While this is lower than the 27% APR, it is still a high interest rate.
3. Interest on credit deposit: While depositing some amount as a credit deposit may earn you a 6% interest, it may not be the best use of your funds as you will still be charged interest on the outstanding credit card balance.
4. Credit utilization: By paying off the credit card balance in full, your credit utilization ratio will be low, which is beneficial for your credit score. A low credit utilization ratio signals to lenders that you are a responsible borrower.

Overall, paying off the full credit card balance before the end of the billing cycle is the best option to avoid interest charges and maintain a healthy credit score.

Figure 2: ChatGPT Response (Q11-AC11).

- **Mathematical Errors:** We consider errors like rounding errors, calculation errors, etc. as mathematical errors.

- **Perceptual Errors:** When the agent misinterprets information given by the user or makes assumptions on un-

known data, we consider these as perceptual errors.

- **Grammatical Errors:** We consider typos, grammatical errors, etc. as grammatical errors (we encountered these errors only in Telugu text generated by ChatGPT).

- **Lack of Visual Aids:** When the agent doesn't use visual aids like tables, graphs, etc. in its response, we consider these as lack of visual aids.

Table 3 shows the percentage of queries for which the chatbots exhibited each of these errors. We also list out the individual query identifiers. Qi denotes the query identifier as previously defined (and also shown in Table 1). ABi and ACi refer to the corresponding Bard and ChatGPT responses respectively. 'i' denotes the identifier (number). Figures 1 and 2 show the response generated by Bard and ChatGPT chatbots respectively. For this one query, Bard made use of a table (though it misinterpreted user information) and ChatGPT did not.

Error Category	Queries	% of Bard Queries	% of ChatGPT Queries
Lack of Personalized Recommendations	Q1-AB1, Q3-AB3, Q3-AC3, Q4-AB4, Q5-AB5, Q6-AC6, Q7-AC7, Q8-AB8, Q9-AB9, Q10-AC10, Q11-AC11, Q12-AB12, Q12-AC12, Q13-AB13	53.84%	46.15%
Mathematical Errors	Q2-AB2, Q9-AC9, Q10-AB10	15.38%	7.69%
Perceptual Errors	Q8-AC8, Q10-AB10, Q11-AB11	15.38%	7.69%
Grammatical Errors	Q6-AC6, Q7-AC7	0%	15.38%*
Lack of Visual Aids	All except Q11-AB11	92.30%	100%

Table 3: Table showing % of queries for which the chatbots exhibited different errors along with individual query-response identifiers. 'Qi' denotes the query identifier, 'ABi' and 'ACi' represent the corresponding Bard and ChatGPT responses respectively where 'i' is the identifier.

Discussion and Conclusion

The application of language models in the finance industry has witnessed a surge in recent times due to their ability to process vast volumes of unstructured data and extract valuable insights. This paper delves into the performance of two prominent language models, Bard and ChatGPT, within the finance domain.

We also find the following challenges in evaluating LLM-based systems for finance domains:

- C1: Changing nature of answers for the same question. How does one create reference test cases since the answers change over time?
- C2: Inability of the chatbots to do numeric reasoning
- C3: Presenting results with easy to follow graphics.
- C4: Support for languages used by customers from different population groups. We considered AAVE - (African American Vernacular English) and Telugu, an Indian language spoken by nearly 100m people world-wide.
- C5: Evaluation the response of users from a diverse set of background. We only considered college students in this study.

C1 can be mitigated by carefully cataloging questions and system answers by identifiers that account for changing behavior over time. For C2, integration with numeric solvers like Wolfram may help (Wolfram 2023) although this makes the systems non-learnable over time. For C3, different data presentation strategies need to be tried. For C4, the LLM models or the chatbots need to be enhanced. For C5, more experiments are needed with inputs carefully modeling the characteristics of the different user groups. These are just preliminary challenges and we expect them to grow as more researchers will try LLM-based systems in complex and diverse application scenarios.

While our study only comprised thirteen queries, we meticulously selected them to cover various categories of credit card finance. However, there exists ample scope for more extensive testing of these chatbots by expanding the number of queries under each category or including additional categories like student loans and stock purchases. By doing so, we can gain a better understanding of the efficacy of language models in different financial domains and improve their functionality in real-world scenarios.

References

Alberto Pozanco, D. B., Kassiani Papasotiriou. 2022. PFPT: a Personal Finance Planning Tool by means of Heuristic Search and Automated Planning.

Bang, Y.; Cahyawijaya, S.; Lee, N.; Dai, W.; Su, D.; Wilie, B.; Lovenia, H.; Ji, Z.; Yu, T.; Chung, W.; et al. 2023. A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023*.

Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Lee, Y. T.; Li, Y.; Lundberg, S.; et al. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Frieder, S.; Pinchetti, L.; Griffiths, R.-R.; Salvatori, T.; Lukasiewicz, T.; Petersen, P. C.; Chevalier, A.; and Berner, J. 2023. Mathematical capabilities of chatgpt. *arXiv preprint arXiv:2301.13867*.

Google, . 2023. Google BARD. In *https://bard.google.com/*.

Hu, B.; Xia, J.; Zheng, J.; Tan, C.; Huang, Y.; Xu, Y.; and Li, S. Z. 2022. Protein Language Models and Structure Prediction: Connection and Progression. *arXiv preprint arXiv:2211.16742*.

Li, X. L.; Kuncoro, A.; Hoffmann, J.; de Masson d'Autume, C.; Blunsom, P.; and Nematzadeh, A. 2022. A systematic investigation of commonsense knowledge in large language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 11838–11855.

Min, B.; Ross, H.; Sulem, E.; Veyseh, A. P. B.; Nguyen, T. H.; Sainz, O.; Agirre, E.; Heinz, I.; and Roth, D. 2021. Recent advances in natural language processing via large pre-trained language models: A survey. *arXiv preprint arXiv:2111.01243*.

OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774.

Wolfram, S. 2023. ChatGPT Gets Its "Wolfram Superpowers"! *https://writings.stephenwolfram.com/2023/03/chatgpt-gets-its-wolfram-superpowers/*.

Wu, S.; Irsoy, O.; Lu, S.; Dabravolski, V.; Dredze, M.; Gehrmann, S.; Kambadur, P.; Rosenberg, D.; and Mann, G. 2023. BloombergGPT: A Large Language Model for Finance. arXiv:2303.17564.

Xiao, Z.; Liao, Q. V.; Zhou, M.; Grandison, T.; and Li, Y. 2023. Powering an AI Chatbot with Expert Sourcing to Support Credible Health Information Access. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 2–18.

Yue, T.; and Au, C. C. 2023. GPTQuant's Conversational AI: Simplifying Investment Research for All. *Available at SSRN 4380516*.

Zhao, W. X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z.; Du, Y.; Yang, C.; Chen, Y.; Chen, Z.; Jiang, J.; Ren, R.; Li, Y.; Tang, X.; Liu, Z.; Liu, P.; Nie, J.-Y.; and Wen, J.-R. 2023. A Survey of Large Language Models. arXiv:2303.18223.