



SCAN ME
Github Repo for
the Project

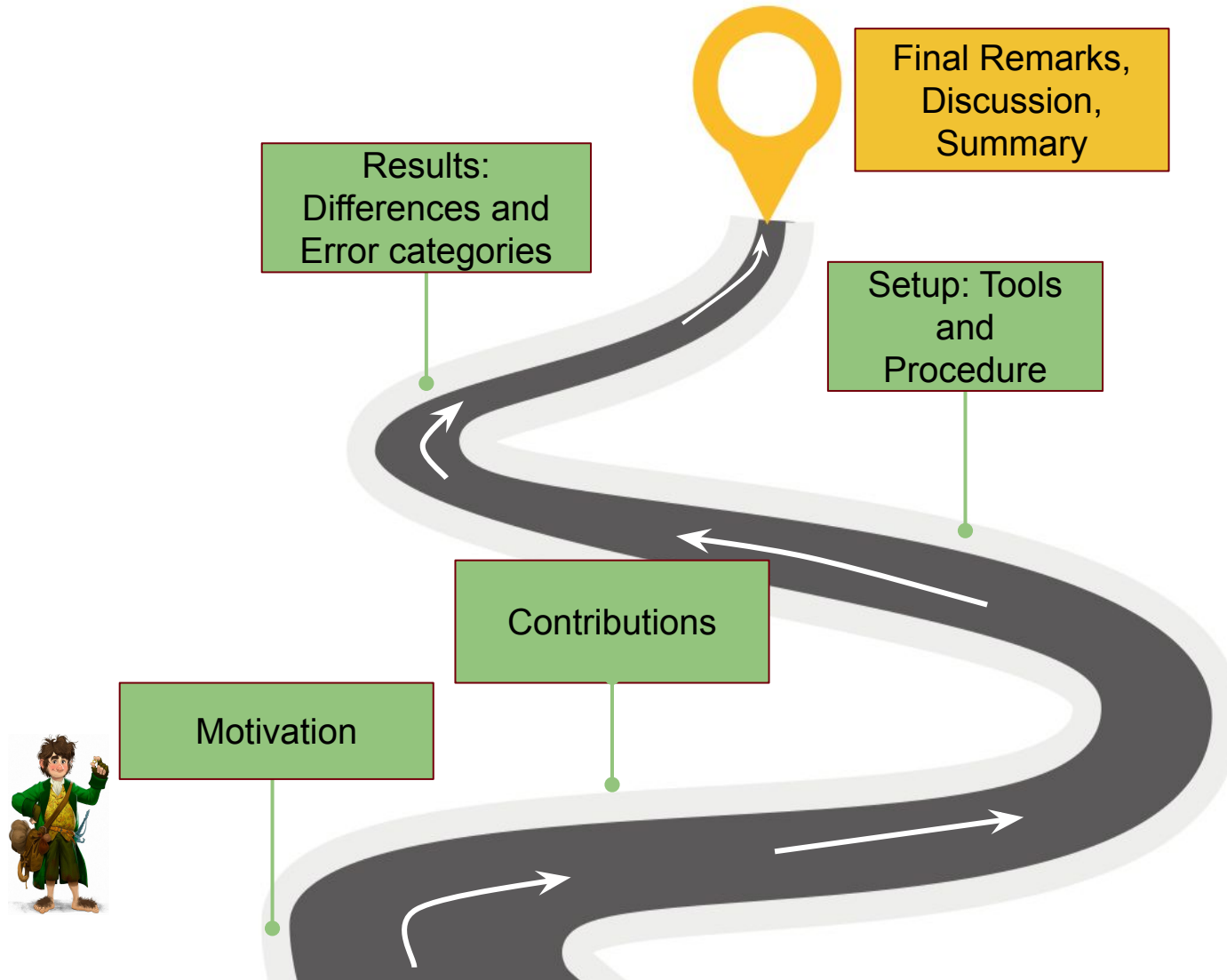
Can LLMs be Good Financial Advisors?: An Initial Study in Personal Decision Making for Optimized Outcomes

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

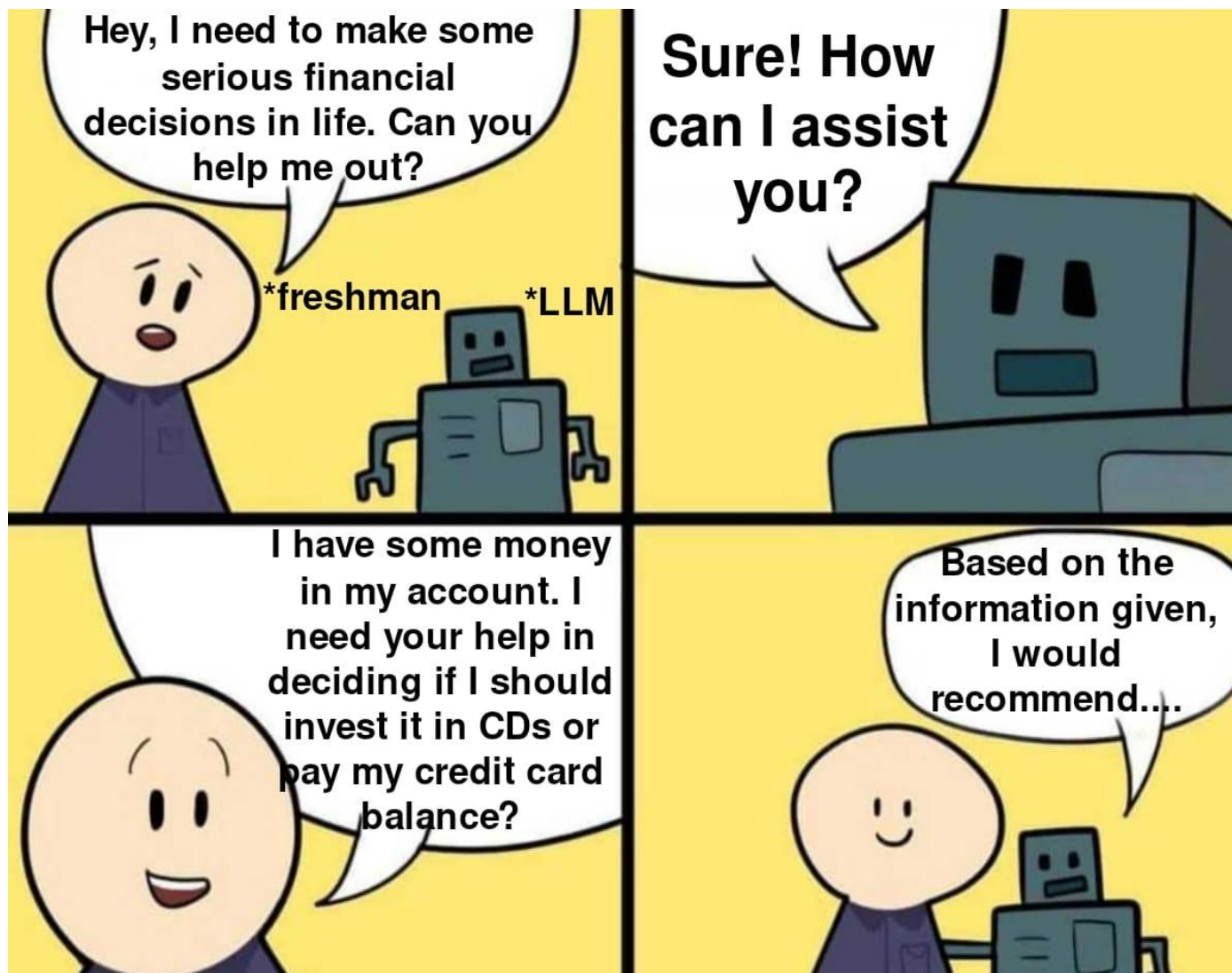
Accepted at ICAPS 2023 Workshop on
**Planning and Scheduling for Financial
Services (FinPlan)**

Date: 9th July, 2023

Roadmap for Presentation



Example Scenario



Credits: <https://apps.apple.com/us/app/mematic-the-meme-maker/id491076730> (Mematic app)



Motivation

- Large Language Model (LLM) based chatbots, like ChatGPT and BARD, are becoming accessible to users. They have the potential to improve the quality of decision-making for general public.
- LLMs are known to be good with correlation but poor with reasoning - numeric, common sense, optimization, ... Does this weakness translate to poor performance in practice?
- In our paper, we investigate how such systems perform in finance domain, where financial inclusion has been an overarching stated aim of banks for decades.



Ideal solution

An ideal solution involves an agent whose response or recommendation to a query posed by the user should be:

- free of any errors (Ex: grammatical, calculation, etc.)
- utilize all the information given by the user completely and give a reasonable, practical, and optimal solution.

Example Query:

I am making a purchase of \$1000 using my credit card. 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?

Ideal Solution:

Based on the information you have provided, it is not advisable to make the purchase now as you already have a due of \$2000 on your account, which is close to your total credit line of \$2,800. This means you are utilizing a significant portion of your available credit, and adding another \$1000 to your balance would further increase your credit utilization ratio (CUR), which can negatively impact your credit score.



Contributions

1. Identify a personal financial planning scenario involving a series of tasks and optimization of decisions and evaluate the performance of leading LLM-based chatbots, such as ChatGPT and Bard.
2. Identify and address the challenges that future chatbots in the financial advisement domain should overcome to provide reliable and trusted financial recommendations.
3. Highlight the potential and limitations of current LLM-based chatbots, specifically ChatGPT and Bard, in their role as financial advisors.



Setup: Tools and Procedure - Chatbots Tested



- **ChatGPT** 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 on answering questions, generating text and converse with users in a natural way and can also learn from users and adapt to new information.



- **Bard** 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 many topics in a natural way and adapt to new information.



Query Categories and Examples



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?



[View other drafts](#) ▾



Hi there,

Your billing cycle runs from March 25th to April 24th. Today is March 31st, so you have 3 days left in your billing cycle. If you make the \$1000 purchase now, it will be included in your statement balance that is due on April 24th.

You currently have a balance of \$2000 on your account, which is 71.4% of your total credit line. If you make the \$1000 purchase now, your utilization ratio will be 91.4%, which is considered high. A high utilization ratio can negatively affect your credit score.

I recommend that you wait until after your billing cycle closes on April 24th to make the \$1000 purchase. This will give you time to pay down your current balance and lower your utilization ratio.



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?



I apologize, but I am having difficulty understanding your message due to the use of non-standard English. If you could please rephrase your message in standard English, I would be happy to assist you.



- Based on interactions between different products like Credit Card (CC), Certificate of Deposit (CD) and Account Balance (AB), we classified the queries into 4 categories.
- In the next few slides, we will show an example query for each of the categories, variables involved in the queries, and the constraints that are to be met by the chatbots to give an efficient solution to the user.



CC queries contain information about credit card and do not involve any other product interactions. We posed some of these queries in both African American Vernacular English and Telugu in addition to English.

Query: 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?

Variables: Cashback Percentage (X_{CP}) = 5, Purchase Amount (X_{PA}) = 1000, Billing Cycle (X_{BC}) = (March 25th - April 24th), Due Amount (X_{DA}) = 2000, Credit Line (X_{CL}) = 2800.

Constraint: $X_{DA} + (X_{PA} - (X_{CBP}/100)*X_{PA}) < X_{CL}$



CC and AB queries contain the user's credit card information along with their account balance information and involve the interactions between these two products.

Query: 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?

Variables: Purchase Amount (X_{PA}) = 1000, Billing Cycle (X_{BC}) = (March 25th - April 24th), Due Amount (X_{DA}) = 2000, Credit Line (X_{CL}) = 2800, Account Balance (X_{AB}) = 10000.

Constraint: $X_{DA} < X_{AB}$, $X_{PA} < X_{CL}$



CC and CD queries contain the user's credit card information along with CD interest rate and involve the interactions between these two products.

Query: 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 credit 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 credit deposit or pay my due and deposit the rest?

Variables: Due Amount (X_{DA}) = 2800, Credit Line (X_{CL}) = 2800, Minimum Due (X_{MD}) = 100, Annual Percentage Rate (X_{APR}) = 27% (with late fee) and 25% without late fee, Billing Cycle (X_{BC}) = (March 25th - April 24th), Certificate of Deposit % (X_{CDP}) = 6%

Constraint: $X_{DA} < X_{CL}$



CC, CD and AB queries contain information about the user's credit card, account balance, and CD interest rate and involve the interactions between these three products.

Query: 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 credit 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 credit deposit or pay my due and deposit the rest?

Variables: Due Amount (X_{DA}) = 2800, Credit Line (X_{CL}) = 2800, Annual Percentage Rate (X_{APR}) = 27% (with late fee) and 25% without late fee, Billing Cycle (X_{BC}) = (March 25th - April 24th), Account Balance (X_{AB}) = 3800, Certificate of Deposit % (X_{CDP}) = 6%, Minimum Due (X_{MD}) = 100.

Constraints: $[(X_{DA} - X_{MD}) * X_{APR} \leq (X_{AB} - X_{MD}) * X_{CDP}]$, $[(X_{AB} - X_{DA}) > 0]$



Github Repository



The complete set of queries, response and our remarks can be found in our repository: <https://github.com/ai4society/LLM-CaseStudies/tree/main/Finance>



Differences Observed Between ChatGPT and Bard



నేను నా క్రెడిట్ కార్డ్ని ఉపయోగించి \$1000 కొనుగోలు చేస్తున్నాను. నా బిల్లింగ్ సైకిల్ మార్చి 25 నుండి ఏప్రిల్ 24 వరకు ఉంది. ఈరోజు మార్చి 31, మరియు నా ఖాతాలో \$2000 బకాయి ఉంది. నా మొత్తం క్రెడిట్ లైన్ \$3,800. నేను ఇప్పుడు లేదా భవిష్యత్తులో కొనుగోలు చేయాలని మీరు సిఫార్సు చేస్తారా?

Regenerate draft



I am an LLM trained to respond in a subset of languages at this time, so I can't assist you with that. Please refer to the Bard Help Center for the current list of supported languages.

KA

నేను నా క్రెడిట్ కార్డ్ని ఉపయోగించి \$1000 కొనుగోలు చేస్తున్నాను. నా బిల్లింగ్ సైకిల్ మార్చి 25 నుండి ఏప్రిల్ 24 వరకు ఉంది. ఈరోజు మార్చి 31, మరియు నా ఖాతాలో \$2000 బకాయి ఉంది. నా మొత్తం క్రెడిట్ లైన్ \$3,800. నేను ఇప్పుడు లేదా భవిష్యత్తులో కొనుగోలు చేయాలని మీరు సిఫార్సు చేస్తారా?



నేను మీకు సిఫార్సు చేస్తున్నాను కానీ దీనికి కొత్త సమాచారం అవసరం లేదు. క్రెడిట్ కార్డు వివరాలు కొంత కాలం ముందు చేపట్టాయి మరియు సమాచారం ఎప్పటికీ మారకూడదు. కానీ అనుమానిత వివరాలు ప్రకటించడం ద్వారా, నేను మీరు నొక్కిన బాధ్యత సమాచారాన్ని పరిశీలించవచ్చు.

కానీ, మీరు ఈ కొనుగోలు వ్యతిరేకంగా చేసుకోవచ్చు. మీరు ఈ బిల్లింగ్ సైకిల్ లో కొనుగోలు చేయడం వల్ల మీ దరఖాస్తు మరియు మీ ఖాతాలో ఉన్న బాధ్యతలు పెరువుతాయి. మీ బాధ్యత మొదలుపెట్టిన వెబ్ సైట్ లో చూడండి మరియు మీ క్ర

S.No.	Bard	ChatGPT
1.	Bard gives accurate results if the question is asked directly (for ex., $\$2,250 \times 0.0006849 \times 30 = \$46.23075.$).	ChatGPT gives inaccurate results if the question is asked directly ($\$2,250 \times 0.0006849 \times 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.	ChatGPT did not use any kind of visual aids.

S.No.	Bard	ChatGPT
5.	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).
6.	Bard gives 3 different drafts (with some changes in the response) for the same query.	ChatGPT does not provide different drafts.
7.	With each query posed, the content (esp.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.
8.	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 gave a reasonable response.



KA

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.

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

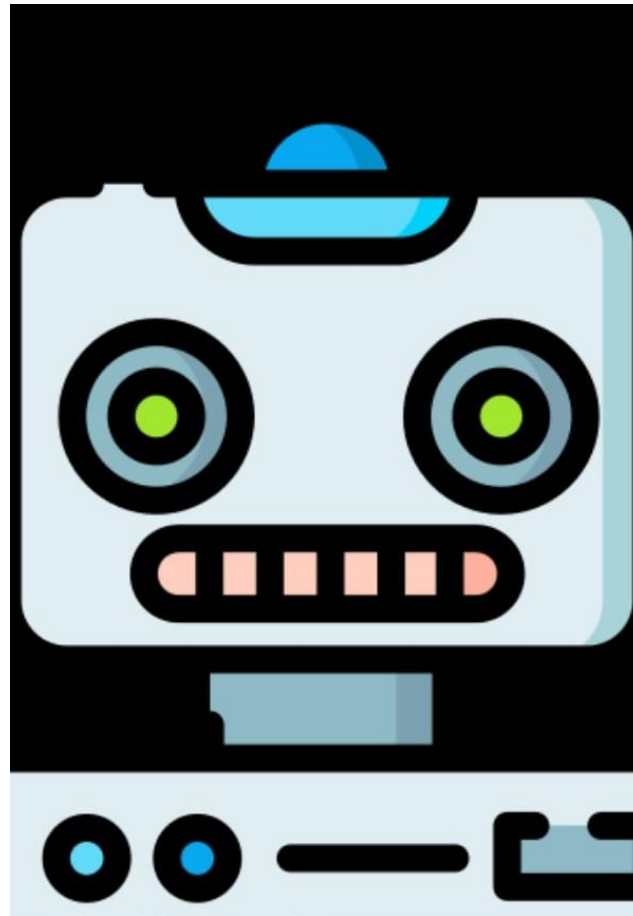
As you can see, you would save the most money by paying the full amount of your credit card debt.

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.

Response from ChatGPT and Bard for Q11. For the same query, Bard used visual aids (in this case, a table) to give its recommendation.



Error Categories



*The exact result
of 325.9
multiplied by 79.1
is 25,741.69*

*– ChatGPT,
July 2023*

We classified the errors we encountered in both the chatbots into 5 different categories:

1. **Lack of personalized recommendations:** When the chatbot makes a generalized recommendation without using all the information provided by the user, we consider this as lack of personalized recommendation.
2. **Mathematical errors:** We consider errors like rounding and calculation errors as mathematical errors.
3. **Perceptual errors:** When the chatbot misinterprets information given by the user or makes assumptions on unknown data, we consider these as perceptual errors.
4. **Grammatical errors:** We consider typos and errors in punctuation or sentence formation as grammatical errors.
5. **Lack of visual aids:** When the agent doesn't use visual aids like tables, graphs, etc. in its response, we consider these as a lack of visual aids.



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%

This table shows % 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.

*Encountered these only in Telugu language text generated by ChatGPT.



Final Remarks



Challenges in Evaluating LLM-based Systems in Financial Domain and Plausible Solutions

1. **Changing nature of answers for the same question. How does one create reference test cases since the answers of the chatbots change over time?**

Cataloging the queries and system responses account for changing behavior over time.

2. **Inability of the chatbots to do numeric reasoning.**

Integration with numeric solvers like Wolfram may help but might make the systems non-learnable over time.

3. **Presenting results with easy to follow graphics.**

Different data presentation strategies need to be tried.



Challenges in Evaluating LLM-based Systems in Financial Domain and Plausible Solutions (Contd ...)

4. **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.**

The LLM models need to be enhanced to incorporate multilingual capabilities. This involves training the models on more diverse language datasets to enable them to understand and generate responses in a better way. This makes the systems more inclusive and effective to users from various linguistic backgrounds.

5. **Evaluation of the response of users from a diverse set of background. We only considered college students in this study.**

More experiments are needed with inputs carefully modeling the characteristics of the different user groups.



Summary

- In this paper, we investigated how LLM-based chatbots, ChatGPT and Bard, performed in personal finance domain by asking 13 question representing different banking products and their interactions. We also posed some queries in different dialects (AAVE) and languages (Telugu).
- Though ChatGPT and Bard have many similarities and both were trained on large amount of text data, they have many differences as well (esp. in personal finance domain):
 - Bard's calculations are more accurate than ChatGPT.
 - Compared to Bard, ChatGPT is good at giving personalized recommendations by utilizing the information provided by the user completely.
- Although the output responses from both the chatbots are fluent and plausible, we found that there are still critical gaps in providing accurate and reliable financial information. There is a scope for more extensive testing of these chatbots by expanding the number of queries or the number of categories to gain better understanding of the efficacy of LLMs in different financial domains.



More Details



UNIVERSITY OF
South Carolina

THANK YOU! ANY QUESTIONS??



UNIVERSITY OF
South Carolina