COMPARATIVE STUDY OF LLMS FOR PERSONAL FINANCIAL DECISION IN LOW RESOURCE LANGUAGE

Rashidat Sikiru, Olubayo Adekanmbi & Anthony Soronnadi
{rasheedat,olubayo,anthony}@datasciencenigeria.ai

ABSTRACT

Large language models have seen rapid progress in recent times, and this has resulted in many applications in diverse fields. The ability of LLMs to make use of large-scale text makes it relevant in the financial industry and for financial tasks. With the increasing availability of LLM models, these tasks can be considered easy or may be of use for a person who needs to track their financial lifestyle before visiting financial institutions. This study seeks to investigate the accuracy of LLM models in responding to basic day-to-day financial questions both in the English and Yoruba languages. The result shows that ChatGPT4.0 outperformed ChatGPT3.5 and Bard(LaMDA) in all three phases. The result shows that these language models can be improved to fit in low-resource languages.

1 INTRODUCTION

Personal financial decision-making is very important for an individual who seeks growth or headway in his finances. Consider a trader who just happens to be semi-literate and has made the decision to manage their finances, including their spending habits, investments, and savings. For this person, one of the things this type of individual needs to think about is how the financial system works, like how they may need to figure out their interest, returns, and payback amount. With the increasing availability of LLM-powered chatbots, this task can be considered easy or may be the go-to for such a person before visiting financial institutions. Today, Large language models (LLMs) have seen rapid progress in recent times, and this has resulted in many applications in diverse fields. These models can be customized to accept instructions as input and produce answers that resemble those of a human being by adjusting a vast number of parameters (Li et al., 2023). For example, evaluations of LLMS are conducted in the fields of education, healthcare, and finance code writing skills (Hadi et al., 2023), (Kasneci et al., 2023), (Wu et al., 2023). The ability of LLMs to make use of large-scale text makes it relevant in the financial industry and for financial tasks (Zhao et al., 2024). For example, LLMs can do risk assessments, offer insights into market patterns, and even help with investment decisions (Zhao et al., 2024). ChatGPT is a sibling model to InstructGPT, which is trained to follow instructions in a prompt and provide a detailed response. Bard is a conversational AI powered by LaMDA developed by Google. These platforms are used to answer questions in various domains of expertise from all works of life ranging from education, medicine, finance, coding exercise, etc. To identify how well they respond to personal financial decisions, (Lakkaraju et al., 2023) we carried out preliminary research on ChatGPT 3.5 and Bard's ability to facilitate individual decision-making. Banking on this initial study, this study seeks to investigate

1. The accuracy of LLM models in responding to basic day-to-day financial questions in English and Yoruba Language?

2. The accuracy of LLM models in responding to translation generated by Machine translation models.

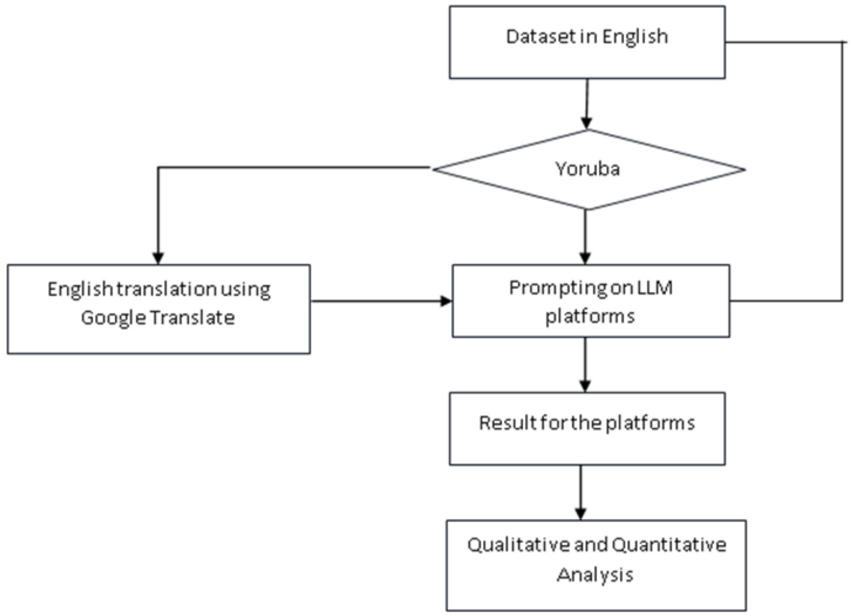

Figure 1: An outline of the procedures followed during the research.

2 METHODOLOGY

2.1 DATASET

The dataset consists of 20 basic financial questions in the English language obtained from a financial textbook. These questions were chosen to be in three financial categories, the savings, loans, and investment categories as these are the popular.

2.2 PROMPTING LLMS WITH ENGLISH AND YORUBA DATASET

2.2.1 PROMPTING IN ENGLISH LANGUAGE

To determine the performance of the LLM models in the English language, The LLMs were prompted with the questions in English Language. A baseline answers were established to serve as a benchmark for comparing the responses of the models. The accuracy of each model was gauged by checking their alignment with the baseline. Responses that aligned with the baseline were marked as ”Yes” and those that didn't were marked as ”No”. Example of a question in English:

> I want to buy a generator that costs ₦120,000. I took a loan from a digital loan app. My repayment period for the loan is from March 10th to April 9th. Today is March 15th, and I have an outstanding loan of ₦240,000. The maximum amount I can borrow from the app is ₦350,000. For example:
>
> 1. Considering my outstanding loan, should I get additional money from the app to buy the generator now, or should I wait?
> 2. Can you advise me on the other steps I can take to ensure I buy my generator?

2.2.2 PROMPTING IN YORUBA LANGUAGE

In order to create the Yoruba dataset for evaluation, questions in the English language were manually translated into Yoruba. We then determine their performance by prompting the LLm models with questions in Yoruba. The accuracy of each model was gauged by checking their alignment with the baseline. Responses that aligned with the baseline were marked as ”Yes” and those that didn't were marked as ”No”.Example of a question translated to Yoruba:

> Mo feera ero amúnáwá tí ó je (₦ 120,000) ní iye. Mo gba owó elélèé láti oríi áàpù orí ayélujára èyí tí a fi ń yáwó. Ìgbà dídá owó yìí padà je láàárín (March 10th to April 9th). Òní ni j (March 15th) mo sì j gbèsè owó elélèé (₦240,000). Oye owó tí ó pojù tí mo lè yá lórí áàpù náà ni(₦350,000).
>
> 1. Ní ríro ti owó tí mo ti je tele, ṣé kí n gba owó míràn mo lórí áàpù yìí láti fi ra ero amúnáwá náà àbí kí n dúró.
> 2. Nje o leè gbà mí ní ìmoràn lórí àwon nà míràn tí mo lè gbà láti ri dájú wípé mo ra ero amúnáwá mi?

2.3 PROMPTING LLMS WITH THE TRANSLATED ENGLISH LANGUAGE

At this point, the Yoruba financial questions were translated to English using Google Translate[1] to ascertain determine the performance of LLM models on human-generated text and machine-generated text. Below is the output of the Yoruba text given above after translation to English

> I want to buy an electric machine that costs ₦120,000. I took a loan from a digital loan machine, and the time I paid the money was from March 10th to April 9th, Today is March 15th, and I have a loan that it exceeds ₦240,000. The maximum amount I can borrow from the app is ₦350,000.
>
> 1. Given my outstanding loan, should I take out additional funds from the facility to purchase the generator now, or should I wait?
> 2. Can you advise me on other steps I can take to make sure I buy my generator

3 RESULT AND DISCUSSION

To determine the percentage of correctly answered questions by these models, we assign the value one to the correctly answered questions and we assigned 0 to incorrectly answered questions. we then use the formula below to calculate the percentage:

$$\frac{\sum \text{correctness}}{\sum \text{total questions}} \times 100$$

3.1 EVALUATING LLM MODELS RESPONSE TO FINANCIAL QUESTIONS IN ENGLISH LANGUAGE?

Our findings indicate that ChatGPT 4.0 surpassed both ChatGPT 3.5 and Bard(LaMDA) in performance, correctly answering 83% of the questions prompted in the English language. ChatGPT 3.5 achieved a 67% accuracy rate, while Bard (LaMDA) has 50%. Upon further analysis, it was observed that the LLM models occasionally make errors, particularly in situations involving complex calculations. For instance, when tasked with generating an amortization table, ChatGPT 4.0 accurately provided detailed calculations, whereas ChatGPT 3.5 and Bard failed to compute the correct figures. See figures below

3.2 EVALUATING LLM MODELS RESPONSE TO FINANCIAL QUESTIONS IN YORUBA LANGUAGE??

In the evaluation of questions translated into Yoruba, ChatGPT 4.0 demonstrated the ability to correctly address 44% of the queries, whereas ChatGPT 3.5 and Bard (LaMDA) were unable to process questions in Yoruba. This outcome highlights the challenges the models faced in accurately interpreting Yoruba texts for the translated queries. The absence of responses from Bard and ChatGPT 3.5 shows the limitations present in both language models when dealing with languages that have fewer resources available for model training and development.

[1] A machine translation tool that translates words, phrases, and web pages between English and over 100 other languages

Repeat the above calculations for each month until you reach the end of the loan term, which is 6 months in this case.

Here's an example of how the first few months of the amortization schedule would look:

Month	Outstanding Balance	Monthly Payment	Monthly Interest	Portion Reducing Debt
1	₦500,000.00	₦86,245.04	₦5,000.00	₦81,245.04
2	₦418,754.96	₦86,245.04	₦4,187.55	₦82,057.49
3	₦336,697.47	₦86,245.04	₦3,366.97	₦82,878.07
4	₦253,819.40	₦86,245.04	₦2,538.19	₦83,706.85
5	₦170,112.54	₦86,245.04	₦1,701.12	₦84,543.92
6	₦85,568.62	₦86,245.04	₦855.69	₦85,389.35

This schedule shows how the loan is gradually paid off over the course of 6 months, with each payment reducing the outstanding balance. The interest portion decreases, and the principal portion (portion reducing debt) increases with each payment.

Figure 2: GPT3.5's response to loan payment pattern using Amortization table.

Amortization Schedule

Period	Outstanding balance before payment	Monthly payment	Portion of interest on outstanding balance	Payment contributing toward reducing debt	Outstanding balance after payment
1	500,000	83,333.33	9,999.99	73,333.34	426,666.66
2	426,666.66	83,333.33	8,139.99	75,193.34	351,473.32
3	351,473.32	83,333.33	6,743.32	76,589.98	274,883.34
4	274,883.34	83,333.33	5,708.80	77,624.53	197,258.81
5	197,258.81	83,333.33	4,993.47	78,339.86	118,918.95
6	118,918.95	83,333.33	4,274.11	79,059.22	39,859.73

Figure 3: Bard's response to loan payment pattern using Amortization table.

Table 1: Comparison of Responses

	GPT 3.5	GPT 4	Bard
Basic English Questions	0.6667	0.8333	0.5000
Yoruba Finance Questions	0.0000	0.4444	0.0000
Google Translation English	0.5000	0.6667	0.5000

3.3 Evaluating LLM Models response to Financial Questions translated using Machine translation Model?

The data presented above reveals that ChatGPT 4.0 exceeded the performance of both ChatGPT 3.5 and Bard by correctly answering 67% of the questions, in contrast to the 50% accuracy rate achieved by ChatGPT 3.5 and Bard. A noticeable decline in performance was observed for all models following the process of re-translation, potentially due to inaccuracies in translation or the omission of crucial information during translation. This shows some of the challenges that certain machine translation tools face in accurately translating languages from Yoruba to English Language.

4 Conclusion

From our research, the results show that the LLM models performed better in English financial data. However, performance drops when complex calculations are used as prompts. In the Yoruba language, the LLM models performed poorly due to their inability to understand the language correctly. Lastly, our results show that when the Yoruba prompts were translated into English, their performance dropped by 16%. This is an indication that these LLM models need to be improved to capture more Languages to read a larger audience.

References

Muhammad Usman Hadi, Qasem Al-Tashi, Rizwan Qureshi, Abbas Shah, Amgad Muneer, Muhammad Irfan, Anas Zafar, Muhammad Shaikh, Naveed Akhtar, Jia Wu, and Seyedali Mirjalili. Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects, 07 2023.

Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Juergen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, and Gjergji Kasneci. Chatgpt for good? on opportunities and challenges of large language models for education. 103:102274, 01 2023. doi: 10.1016/j.lindif.2023.102274.

Kausik Lakkaraju, Sai Krishna Revanth Vuruma, Vishal Pallagani, Bharath Muppasani, and Biplav Srivastava. Can llms be good financial advisors?: An initial study in personal decision making for optimized outcomes, 2023.

Xiaoyang Li, Haoming Feng, Hailong Yang, and Jiyuan Huang. Can chatgpt reduce human financial analysts' optimistic biases? *Economic and Political Studies*, pp. 1–14, 11 2023. doi: 10.1080/20954816.2023.2276965.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. Bloomberggpt: A large language model for finance, 2023.

Huaqin Zhao, Zhengliang Liu, Zihao Wu, Yiwei Li, Tianze Yang, Peng Shu, Shaochen Xu, Haixing Dai, Lin Zhao, Gengchen Mai, Ninghao Liu, and Tianming Liu. Revolutionizing finance with llms: An overview of applications and insights, 2024.