

Comparative Analysis of Decision-Making Efficiency of Large Language Models

Mirza Niaz Zaman Elin

Founder, MedTheme Corporation

Abstract:

Large language models (LLMs) have emerged as powerful tools in the field of artificial intelligence (AI), attracting considerable attention from researchers and practitioners. These models demonstrate remarkable capabilities in various tasks, including decision-making. This paper aims to compare the decision-making efficiency of two prominent LLMs, Bard and GPT, across different domains. To conduct a comprehensive evaluation, a set of carefully designed questions was used to assess the performance of Bard and GPT in specific decision-making contexts. Through quantitative analysis, we aimed to quantify their abilities and identify potential variations in their performance. The results of our study revealed interesting insights into the decision-making efficiency of Bard and GPT across different domains. In the domain of logical reasoning and error detection, both Bard and GPT exhibited similar performance, but GPT outperformed Bard in data analysis by a notable margin. This finding suggests that GPT possesses stronger analytical abilities, enabling it to make more accurate and reliable decisions in contexts that require accurate data analysis and interpretation. The comparative analysis of Bard and GPT's decision-making efficiency highlights the significance of considering the specific domains and tasks when evaluating the performance of LLMs. It underscores the fact that different LLMs may possess domain-specific strengths and weaknesses, which can have a profound impact on their decision-making capabilities. Future research endeavors may involve expanding the evaluation to additional domains and considering a larger sample of questions to enhance the reliability and generalizability of the findings. Moreover, investigating the interpretability and explainability of LLMs in decision-making processes could shed further light on their decision-making strategies and enhance trust and transparency in their applications. This paper contributes to the growing body of research on LLMs by comparing the decision-making efficiency of Bard and GPT across different domains. The findings highlight the relative strengths of each model, emphasizing the importance of domain-specific considerations in decision-making tasks. By leveraging the capabilities of LLMs, practitioners can harness their potential to improve decision-making processes in diverse real-world applications.

Keywords: large language models, decision-making, efficiency, Bard, GPT

Introduction

Large language models (LLMs) have revolutionized the field of artificial intelligence (AI) and have garnered significant attention from researchers and practitioners. These models, built on advanced deep learning architectures, are trained on massive datasets to acquire a deep understanding of human language and its nuances. Consequently, LLMs exhibit remarkable capabilities across various AI tasks, including natural language processing, text generation, and even decision-making.

In recent years, decision-making has emerged as a critical area where LLMs have showcased their potential[1]. By leveraging their extensive language understanding and pattern recognition abilities, LLMs can analyze complex information, reason through different scenarios, and generate informed decisions. This has opened up new possibilities for applying LLMs in decision-making processes across diverse domains, ranging from finance and healthcare to customer service and legal analysis.

Among the numerous LLMs developed, Bard and GPT (Generative Pre-trained Transformer) have emerged as prominent contenders. Bard is known for its superior logical reasoning capabilities, while GPT has garnered attention for its exceptional performance in tasks such as text generation and language understanding. Comparing the decision-making efficiency of these two LLMs can provide valuable insights into their relative strengths and weaknesses, enabling practitioners to make informed choices when employing LLMs in decision-making contexts.

The primary objective of this paper is to conduct a comprehensive comparative analysis of Bard and GPT in terms of their decision-making efficiency across different domains. By evaluating their performance using carefully designed domain-specific questions, we aim to quantify their abilities and uncover any discernible variations in their decision-making effectiveness.

Understanding the relative performance of Bard and GPT in decision-making is of paramount importance for several reasons. Firstly, decision-making is a complex cognitive process that involves assessing information, reasoning, and selecting the best course of action. LLMs have the potential to augment and automate decision-making processes, thereby improving efficiency and accuracy. However, the effectiveness of LLMs may vary depending on the specific domain and context in which decisions are made. By comparing Bard and GPT, we can identify the domains where each model excels, allowing us to harness their strengths in relevant decision-making scenarios.

Secondly, real-world decision-making often involves different types of tasks that rely on diverse cognitive skills. For example, logical reasoning tasks require the ability to analyze relationships and draw valid conclusions, while error detection tasks demand a keen eye for identifying deviations and inconsistencies. By examining the performance of Bard and GPT across these varied domains, we can gain insights into their respective strengths and weaknesses, which can guide practitioners in selecting the most suitable model for specific decision-making requirements.

Lastly, as LLMs continue to advance and find applications in critical decision-making domains such as healthcare diagnostics, legal analysis, and financial forecasting, it becomes essential to understand their capabilities and limitations. By conducting a thorough evaluation of Bard and GPT, we aim to contribute to the knowledge base surrounding LLMs and provide valuable insights for practitioners and researchers alike.

In the subsequent sections of this paper, we will detail the methodology employed for evaluating Bard and GPT's decision-making efficiency, present the specific domains chosen for analysis, and describe the designed questions tailored to each domain. We will then present the results of the comparative analysis,

discussing the performance of Bard and GPT in each domain. Finally, we will provide a comprehensive discussion and conclusion, summarizing the findings and outlining the implications for the practical application of LLMs in decision-making contexts.

Through this comparative study, we aim to enhance our understanding of the decision-making efficiency of Bard and GPT, empowering practitioners to make informed decisions regarding the selection and utilization of LLMs in real-world scenarios.

Methods

To conduct a comprehensive comparative analysis of Bard and GPT's decision-making efficiency[2], we designed a rigorous methodology that involved the formulation of domain-specific questions and the evaluation of their performance on these tasks. The following section outlines the steps taken to ensure a fair and systematic comparison between the two LLMs.

1. Selection of Domains:

We carefully selected a diverse set of domains to assess the decision-making capabilities of Bard and GPT. These domains were chosen to represent different cognitive tasks and decision-making contexts, allowing us to evaluate the models' performance across a range of scenarios. The selected domains included logical reasoning dependent decision making (LRDM), error detection dependent decision making (EDM), and data analysis dependent decision making (DADM).

2. Question Design:

For each domain, we designed a set of questions specifically tailored to assess the decision-making efficiency of Bard and GPT. The questions were carefully crafted to target the cognitive skills and abilities required in each domain. In the LRDM domain, the questions focused on logical reasoning, deduction, and inference. In the EDM domain, the questions aimed to evaluate the models' ability to identify errors, inconsistencies, or anomalies. In the DADM domain, the questions were designed to assess the models' proficiency in analyzing and interpreting data to make informed decisions.

3. Presentation of Questions:

Both Bard and GPT were presented with the domain-specific questions in a controlled environment. Each model was given the same set of questions to ensure fairness and eliminate any potential bias arising from differences in the question sets. The questions were presented in a standardized format to ensure consistency across the evaluation process.

4. Scoring and Performance Evaluation:

To quantify the performance of Bard and GPT, we assigned a value of 1 point for each correctly answered question. The total score for each model was calculated based on the number of questions answered correctly within each domain. By adopting this scoring approach, we were able to compare the decision-making efficiency of Bard and GPT in a quantitative manner.

5. Data Collection:

We collected the response data from Bard and GPT for each domain and compiled the results for further analysis. The data included the number of correctly answered questions and the corresponding scores for each LLM in each domain.

6. Statistical Analysis:

To analyze the performance of Bard and GPT, we conducted a statistical analysis of the collected data. This analysis involved calculating accuracy percentages for each model within each domain, allowing us to determine their relative strengths and weaknesses in different decision-making contexts.

By employing this methodology, we ensured a systematic and objective evaluation of Bard and GPT's decision-making efficiency. The design of domain-specific questions and the use of standardized scoring criteria enabled us to assess their performance on various cognitive tasks and provide insights into their relative capabilities in different decision-making domains.

Materials

The materials used in this study primarily consisted of the specially designed questions that were utilized to evaluate the decision-making efficiency of Bard and GPT in different domains. These questions served as the primary means to assess the cognitive abilities and performance of the two LLMs in various decision-making contexts.

The design of these questions was crucial to ensure that they effectively targeted the specific skills and competencies required in each domain. Extensive consideration was given to formulating questions that adequately measured the logical reasoning abilities, error detection capabilities, and data analysis proficiencies of Bard and GPT.

The process of question design involved a thorough review of existing literature and established frameworks for assessing decision-making skills. This allowed us to draw upon established principles and guidelines to create questions that were valid and reliable indicators of the LLMs' decision-making efficiency. The questions were carefully crafted to present realistic scenarios and challenges that mirrored real-world decision-making situations.

To maintain consistency and eliminate potential biases, the same set of questions was presented to both Bard and GPT during the evaluation process. Each question was presented in a standardized format to ensure uniformity in the way the questions were interpreted and answered by the models.

The materials used in this study were essential for providing a standardized and controlled environment for evaluating the decision-making efficiency of Bard and GPT. The questions, specifically designed for each domain, formed the basis for quantitatively assessing the performance of the LLMs and comparing their abilities in different decision-making tasks.

It is worth noting that the study also relied on the computational resources and infrastructure required to run the LLMs and collect their responses to the domain-specific questions. The computational resources

ensured the efficient execution of the evaluation process, enabling us to gather the necessary data for analysis.

Overall, the materials used in this study, primarily consisting of the domain-specific questions, played a vital role in evaluating the decision-making efficiency of Bard and GPT. These materials facilitated a standardized evaluation process, allowing for a fair and objective comparison of the two LLMs' performance in different decision-making domains.

Procedure

To evaluate the decision-making efficiency of Bard and GPT, a systematic and standardized procedure was followed. The following steps were taken to ensure a fair comparison between the two LLMs:

1. Presentation of Domain-Specific Questions:

Both Bard and GPT were presented with a set of domain-specific questions that were cognitive tasks in nature. These questions were carefully designed to assess the decision-making capabilities of the LLMs in different domains, including logical reasoning, error detection, and data analysis. Each question was tailored to target the specific skills and competencies required in its respective domain.

2. Scoring System:

A scoring system was established to assign a value of 1 point to each correctly answered question. This system allowed for a quantifiable measurement of the LLMs' performance in each domain. By assigning a point for every correctly answered question, a numerical score was obtained for both Bard and GPT, reflecting their decision-making proficiency within each domain.

3. Calculation of Total Score:

The total score for each LLM was calculated based on the number of correctly answered questions. By summing up the individual scores obtained for each question within a domain, a cumulative score was determined. This total score provided an overall measure of the decision-making efficiency of Bard and GPT in each evaluated domain.

The procedure ensured a consistent and standardized approach to evaluating the decision-making efficiency of Bard and GPT. By presenting both LLMs with the same set of domain-specific questions and using a scoring system, a clear comparison was made possible. The total scores allowed for an objective assessment of the LLMs' performance, enabling a meaningful comparison of their decision-making abilities in different domains.

It is important to note that the procedure employed in this study focused on quantitatively evaluating the decision-making efficiency of Bard and GPT. While this approach provided valuable insights into their performance, it is also essential to consider qualitative factors and the context in which decision-making occurs. Therefore, the results should be interpreted in conjunction with a broader understanding of the LLMs' capabilities and limitations in real-world decision-making scenarios.

5. Data Analysis

The data analysis focused on three domains of decision-making efficiency: Logical Reasoning Dependent Decision Making (LRDM), Error Detection Dependent Decision Making (EDM), and Data Analysis Dependent Decision Making (DADM). For each domain, a specific set of questions was designed to evaluate the decision-making performance of Bard and GPT.

In the domain of Logical Reasoning (LRDM), Bard and GPT both obtained a score of 3 points, indicating a similar proficiency in logical reasoning tasks.

Moving on to the Error Detection (EDM) domain, both Bard and GPT achieved a score of 2 points, indicating a similar level of effectiveness in error detection tasks. This suggests that both models performed adequately in identifying and detecting errors within the given context.

In the Data Analysis (DADM) domain, Bard obtained a score of 11 points, while GPT scored 17 points. These scores indicate that GPT outperformed Bard in the data analysis tasks. GPT’s higher score suggests a greater proficiency in analyzing and interpreting data to make informed decisions[Table-1].

Table-1: Decision-Making Efficiency Scores by Domain: Bard vs. GPT

Domain	Total Questions	Bard Correct	GPT Correct	Accuracy - Bard	Accuracy - GPT
Logical Reasoning (LRDM)	4	3	3	75%	75%
Error Detection (EDM)	2	2	2	100%	100%
Data Analysis (DADM)	19	11	17	57.89%	89.47%

Discussion

Decision-making is a multifaceted process influenced by numerous factors beyond the specific domains evaluated in this study[3]. Other important considerations include context, subject matter expertise, understanding of nuances, and the ability to integrate information from diverse sources. Therefore, while the findings provide insights into the relative performance of Bard and GPT, they should be interpreted within the broader context of decision-making and the specific requirements of real-world applications[4,5].

The comparative analysis revealed that both Bard and GPT demonstrated similar capabilities in logical reasoning and error detection but GPT demonstrated stronger performance in data analysis and interpretation[6,7].The findings underscore the importance of leveraging the specific capabilities of LLMs and tailoring their use to the requirements of different decision-making domains. By understanding the strengths and weaknesses of LLMs in decision-making, researchers and practitioners can make informed

choices regarding the selection and application of LLMs as decision-making tools in various domains and real-world scenarios.

Conclusion

This comparative study provides valuable insights into the decision-making efficiency of two large language models (LLMs), Bard and GPT, across different domains. The findings highlight the variations in performance between Bard and GPT, emphasizing the importance of considering the specific task requirements and domains when selecting an LLM for decision-making purposes.

Different LLMs may possess domain-specific strengths and weaknesses, and their performance can vary depending on the nature of the decision-making task. Understanding these variations allows researchers and practitioners to make informed choices regarding the selection and application of LLMs in real-world decision-making scenarios.

It is important to note that decision-making is a complex process influenced by various factors beyond the capabilities of LLMs alone. Contextual understanding, subject matter expertise, and human judgment play critical roles in decision-making. LLMs should be viewed as tools that can augment and support decision-making processes, rather than replace human involvement entirely. Therefore, the findings of this study should be interpreted in conjunction with the broader context of decision-making, incorporating human judgment and expertise.

Further research is warranted to delve deeper into the factors contributing to the strengths and weaknesses of LLMs in different decision-making domains. Investigating the integration of domain-specific knowledge and reasoning capabilities into LLMs can potentially enhance their decision-making performance across a wider range of tasks and domains.

By leveraging the specific capabilities of LLMs and considering the specific requirements of decision-making tasks[8], these powerful AI systems can be effectively utilized as valuable decision-making tools in various applications. Continued research and development in this field will pave the way for advancements in AI-assisted decision-making, ultimately benefiting numerous domains and industries.

References

1. Bagnoli, C., Dal Mas, F., & Massaro, M. (2019). The 4th industrial revolution: Business models and evidence from the field. *International Journal of E-Services and Mobile Applications*, 11(3), 34–47. <https://doi.org/10.4018/IJESMA.2019070103>
2. Byerly, S., Maurer, L. R., Mantero, A., Naar, L., An, G., & Kaafarani, H. M. A. (2021). Machine Learning and Artificial Intelligence for Surgical Decision Making. *Surgical Infections*, 22(6), 626–634. <https://doi.org/10.1089/SUR.2021.007>
3. Garousi Mokhtarzadeh, N., Amoozad Mahdiraji, H., Jafari-Sadeghi, V., Soltani, A., & Abbasi Kamardi, A. A. (2020). A product-technology portfolio alignment approach for food industry: a multi-criteria decision making with z-numbers. *British Food Journal*, 122(12), 3947–3967. <https://doi.org/10.1108/BFJ-02-2020-0115/FULL/XML>

4. Grewal, D., Roggeveen, A. L., & Nordfält, J. (2017). The Future of Retailing. *Journal of Retailing*, 93(1), 1–6. <https://doi.org/10.1016/J.JRETAI.2016.12.008>
5. Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586. <https://doi.org/10.1016/J.BUSHOR.2018.03.007>
6. Loftus, T. J., Tighe, P. J., Filiberto, A. C., Efron, P. A., Brakenridge, S. C., Mohr, A. M., Rashidi, P., Upchurch, G. R., & Bihorac, A. (2020). Artificial Intelligence and Surgical Decision-making. *JAMA Surgery*, 155(2), 148–158. <https://doi.org/10.1001/JAMASURG.2019.4917>
7. L'osservatore Romano. (2020). La robotica al servizio del bene comune - L'Osservatore Romano. <https://www.osservatoreromano.va/it/news/2020-11/quo-257/la-robotica-al-servizio-del-bene-comune.html>
8. Loureiro, S. M. C., Guerreiro, J., & Tussyadiah, I. (2021). Artificial intelligence in business: State of the art and future research agenda. *Journal of Business Research*, 129, 911–926. <https://doi.org/10.1016/J.JBUSRES.2020.11.001>