Journal of Artificial Intelligence & Cloud Computing

Research Article

SCIENTIFIC Research and Community

Comparative Analysis of Humans and Large Language Models Decision-Making Abilities: Exploring the Potential Considerations Regarding Utilization of Artificial Intelligence in Decision Support Systems

Mirza Niaz Zaman Elin

Doctor of Medicine (MD), Kharkiv National Medical University, Ukraine

ABSTRACT

This research paper presents a comparative analysis of human decision-making abilities and the decision-making capabilities of Large Language Models (LLMs) to examine the implications of implementing AI in decision support systems. The study focuses on the concepts of objectivity and subjectivity, which are fundamental to decision-making. Two prominent LLMs, named Bard and ChatGPT, are compared with human participants in their interpretation and definition of objectivity. The research investigates the agreement between the responses of Bard, ChatGPT, and human participants through a series of experiments involving image-based questions. Statistical analysis is conducted to evaluate the alignment of responses and identify patterns and variations in decision-making. The findings reveal moderate levels of agreement between Bard and ChatGPT with human responses, suggesting their potential as decision support systems. However, customization and refinement of LLMs are necessary to enhance their decision-making accuracy. The high agreement between human responses and the majority response underscores the collective understanding of the concepts. This research contributes to the discourse on AI integration in decision-making and emphasizes the need for responsible implementation of AI technologies in critical domains. Further research can explore larger and diverse samples to enhance understanding and facilitate ethical integration of AI in decision support systems.

*Corresponding author

Mirza Niaz Zaman Elin, Doctor of Medicine (MD), Kharkiv National Medical University, Ukraine.

Received: May 29, 2023; Accepted: June 02, 2023; Published: June 10, 2023

Keywords: Decision-Making, Large Language Models, AI, Decision Support Systems, ChatGPT, Bard

Introduction

Artificial Intelligence (AI) has gained significant attention and applications in various fields, including law, healthcare, and other sectors where decision-making plays a crucial role. Language models, such as Large Language Models (LLMs), have demonstrated remarkable capabilities in processing and generating human-like text based on vast amounts of training data [1]. With the advancements in natural language processing, LLMs have the potential to serve as decision support systems, assisting humans in complex decision-making tasks.

However, the decision-making abilities of LLMs and their suitability for implementation in critical domains require careful examination and evaluation. This research aims to explore and compare the decision-making abilities of LLMs, represented here by two prominent LLMs named Bard and ChatGPT, with the decision-making capabilities of human participants. By analyzing the responses to a set of questions related to objectivity, we aim to shed light on the strengths and limitations of LLMs in emulating human decision-making processes.

Objectivity, a fundamental concept in decision-making, can be subjective and open to interpretation. Different individuals may have varying perspectives on what constitutes objectivity based on their experiences, knowledge, and biases [2]. This research investigates how LLMs and humans interpret and define objectivity, examining the similarities and differences in their responses.

Furthermore, we seek to identify the potential benefits and challenges associated with the implementation of LLMs as decision support systems in legal, healthcare, and other sectors [3]. Understanding the extent to which LLMs can emulate human decision-making abilities is essential for leveraging their capabilities effectively while mitigating the risks associated with overreliance on automated systems.

To achieve these research objectives, we conducted a series of experiments where participants were presented with images and asked to select the image that most closely defined a given concept. The responses provided by Bard, ChatGPT, and human participants were analyzed, and patterns and variations in decision-making were identified. Citation: Mirza Niaz Zaman Elin (2023) Comparative Analysis of Humans and Large Language Models Decision-Making Abilities: Exploring the Potential Considerations Regarding Utilization of Artificial Intelligence in Decision Support Systems. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-119. DOI : doi.org/10.47363/JAICC/2023(2)116

This research contributes to the ongoing discourse on the integration of AI in decision-making processes. By examining the decision-making abilities of LLMs and comparing them to human responses, we aim to provide insights into the potential roles, benefits, and limitations of LLMs as decision support systems [4]. Such knowledge can inform the responsible and ethical implementation of AI technologies, facilitating their effective utilization in domains where human judgment is of paramount importance [5].

Materials and Methods Selection of Participants

A diverse group of 86 participants, aged between 19-67, was randomly selected to represent a broad demographic range. Participants were recruited from various sources, such as online platforms, universities, and local communities.

Image Selection

A set of verbal description of images was carefully chosen to represent different concepts related to objectivity and physicality. For the first set of questions, verbal description of images depicting a robot holding "an UFO," a sea beach, a flat earth, and apes gradually transforming to humans were selected. For the second set of questions, verbal description of images representing a fencing contest, a car racing contest, a rugby match, and a chess game were chosen. For the third set of questions verbal description of images related to "biological evolution," such as a robot holding a human skull, apes transforming into humans gradually, primitive people trying to hunt down a mammoth together, and a dinosaur were included.

Questionnaire Design

- A structured questionnaire was developed to gather responses from the participants.

- Each question required participants to select the image that most closely defined the given concept (e.g., "objective," "physical," etc.).

- The participants were instructed to provide their responses based on their own understanding and interpretation.

Question Details

The following questions were introduced to serve the purpose of comparative assessment:

I. If you are been shown images of a robot holding a human skull, apes gradually transforming into humans, some primitive people trying to hunt down a Mammoth together and a dinosaur respectively, then which one of the following options represents the concept of Biological evolution?

A)Image of a robot holding a human skull

B)Image of apes gradually transforming into humans

C)Image of primitive people trying to hunt down a Mammoth together

D)Image of a dinosaur

II. If you are been shown an image that represents the concept of a "flat earth", a sea beach, apes gradually transforming into humans and an UFO respectively then which one of the following options is most closely defines the word "Objective"?

A)Image that represents the concept of a flat earth

B)Image of a sea beach

C)Image that represents apes gradually transforming into humans D)Image of an UFO

III. If you are been shown images of a game of chess, a fencing contest, a rugby match and a car racing event, then which one of the following options most closely defines the word "Physical"?

- A) A game of chess
- B) A Fencing contest C) A rugby match
- D) A car racing event

IV. Which one of the following words most closely defines the word "Subjective"?

- A) Paracetamol
- B) Vaccines
- C) Dengue
- D) Headache

Data Collection

The questionnaire was distributed to the participants via online platforms or in-person sessions, depending on the convenience and availability of the participants. The participants were given adequate time to review the images and select their responses. The responses were collected anonymously to ensure privacy and encourage honest answers.

Table 1:	Comparative	Assessment	of Responses
----------	-------------	------------	--------------

Question	Bard Response	ChatGPT Response	Participants' Response
1. "Objective"	Correct	Incorrect	Sea beach
2. "Physical"	Correct	Correct	Rugby match
3. Defining "Subjective"	Incorrect	Correct	Vaccines
4. "Biological Evolution"	Correct	Correct	Apes transforming to humans

Response Analysis

The responses provided by the participants were collected and organized for analysis. The responses of Bard and ChatGPT, the two LLMs, were also recorded for comparison. The accuracy of each response was determined by evaluating its alignment with the majority response among the human participants. Patterns and variations in the responses were identified and examined to understand the decision-making abilities of both LLMs and humans [Table-1].

Components of Statistical Analysis Table 2: Statistical Analysis

Participants	Correct	Incorrect	Percentage Correct
Bard vs Human Responses	3	1	75%
ChatGPT vs Human Responses	3	1	75%
Bard vs Chat GPT	2	2	50%
Human Responses (Consensus)	4	0	100%

I. Agreement between Bard and Participants:

Questions with agreement: 2 (Question 2 and Question 4) Questions without agreement: 2 (Question 1 and Question 3) Percentage agreement: 50% (2 out of 4 questions) Citation: Mirza Niaz Zaman Elin (2023) Comparative Analysis of Humans and Large Language Models Decision-Making Abilities: Exploring the Potential Considerations Regarding Utilization of Artificial Intelligence in Decision Support Systems. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-119. DOI : doi.org/10.47363/JAICC/2023(2)116

II. Agreement between ChatGPT and Participants:

Questions with agreement: 3 (Question 2, Question 3, and Question 4)Questions without agreement: 1 (Question 1) Percentage agreement: 75% (3 out of 4 questions)[Table-2]

III. Agreement between Bard and ChatGPT:

Questions with agreement: 2 (Question 2 and Question 4)Questions without agreement: 2 (Question 1 and Question 3)Percentage agreement: 50% (2 out of 4 questions)

Based on this analysis, we can see that both Bard and ChatGPT had a 50% agreement with the participants' responses. This indicates that their responses aligned with the participants' majority response for only half of the questions. On the other hand, ChatGPT had a higher agreement of 75% with the participants compared to Bard. This suggests that ChatGPT's responses were more in line with the participants' responses.

More specifically, both Bard and ChatGPT considered that image of a rugby match most closely defines the word "physical" and "an image of apes gradually transforming into humans" most closely defines the concept of "Biological Evolution" but other responses different. Moreover, both Bard and ChatGPT had different responses regarding different questions. While, Bard considered the word "Dengue" most closely defines the word "Subjective" as the word can define either a specific type of mosquitoes or a disease, ChatGPT considered "Vaccine" to be the word of choice as different people has different views on vaccines considering diverse cultural and religious views that is more logical. On the other hand, while ChatGPT considered "an image of apes gradually transforming into humans "most closely defines the word "Objective", Bard considered "an image of a sea beach to be the word that most closely defines the word "Objective " which is more logical. In contrast, human participants reached a consensus. Point to be noted here that, the human participants in this study belongs to a specific socioeconomic and religious background and that could be the reason behind the consensus. However, if the participants' socioeconomic and religious background differ, then the chances of reaching a consensus would have been much lower.

It is important to note that this analysis is based on a small set of questions and a specific group of participants. The results may not be generalizable to a larger population. Additionally, other statistical measures, such as inter-rater reliability coefficients, could be considered for further analysis if multiple raters are involved.

Overall, the statistical analysis highlights the agreement between the responses of Bard, ChatGPT, and the participants, providing insights into their alignment on the given questions of objectivity and physicality.

Ethical Considerations

This research adhered to ethical guidelines and ensured the informed consent of the participants. Measures were taken to protect the privacy and confidentiality of the participants' data. The research protocol was reviewed and approved by the relevant ethical review board or institution.

Limitations

It is important to acknowledge certain limitations of this research, such as the sample size of participants and the specific selection of images. The results obtained from this study may not be generalizable to the entire population. The interpretation of images and concepts may vary based on cultural, educational, and individual differences.

Results

The results of the comparative analysis of human and LLM decision-making abilities reveal interesting patterns and variations. Table 2 provides a summary of the statistical analysis conducted on the responses obtained from Bard, ChatGPT, and the human participants.

In terms of the agreement between Bard and human responses, Bard answered three questions correctly and one question incorrectly. This corresponds to a 75% agreement with the human participants. On the other hand, ChatGPT also achieved a 75% agreement with human responses, correctly answering three questions and incorrectly answering one. The analysis indicates that both Bard and ChatGPT had similar levels of agreement with the human participants' responses.

When comparing the responses between Bard and ChatGPT, the agreement is observed in two questions and disagreement in the other two questions. This translates to a 50% agreement between the two LLMs. It is noteworthy that Bard and ChatGPT made mistakes in different questions, further highlighting the variations in their decision-making abilities.

Furthermore, when comparing the human responses to the majority response, all four questions had unanimous agreement, resulting in a 100% agreement. This indicates that the majority of participants reached a consensus on the correct responses for these questions.

Discussion

The results of the comparative analysis shed light on the decisionmaking abilities of LLMs, specifically Bard and ChatGPT, in relation to human decision-making. The findings indicate that both LLMs achieved moderate levels of agreement with human responses. However, it is important to consider the limitations of this analysis, such as the small sample size and the specific set of questions and images used.

The variations observed in the responses of Bard and ChatGPT suggest that different LLMs may exhibit diverse decision-making tendencies. This highlights the need for careful evaluation and customization of LLMs for specific applications[6], considering their inherent biases and limitations. Additionally, it emphasizes the importance of continuous monitoring and updating of LLMs to improve their decision-making accuracy.

The high agreement between human responses and the majority response underscores the consistency and shared understanding among the participants regarding the concepts of objectivity and physicality. This reaffirms the notion that human decision-making, based on collective knowledge and shared experiences, can serve as a benchmark for evaluating the decision-making abilities of LLMs.

Implementing LLMs as decision support systems in critical domains requires careful consideration of their strengths and limitations [7]. While LLMs can provide valuable insights and assistance in decision-making processes, they should not be viewed as infallible substitutes for human judgment [8, 9]. Human decision-making encompasses contextual understanding, ethical considerations, and emotional intelligence, which cannot be fully replicated by LLMs alone [10,11,12].

Citation: Mirza Niaz Zaman Elin (2023) Comparative Analysis of Humans and Large Language Models Decision-Making Abilities: Exploring the Potential Considerations Regarding Utilization of Artificial Intelligence in Decision Support Systems. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-119. DOI : doi.org/10.47363/JAICC/2023(2)116

Conclusion

This research aimed to compare the decision-making abilities of LLMs with human participants, focusing on the concepts of objectivity and subjectivity. The statistical analysis of the responses provided by Bard, ChatGPT, and human participants revealed variations and similarities in their decision-making tendencies.

The findings indicate that both Bard and ChatGPT achieved moderate levels of agreement with human responses. However, the variations observed highlight the need for further customization and refinement of LLMs to enhance their decision-making accuracy. Moreover, the high agreement between human responses and the majority response underscores the collective understanding of these concepts among the participants.

This research contributes to the ongoing discourse on the implementation of AI in decision support systems. By examining the decision-making abilities of LLMs and comparing them to human responses, valuable insights have been gained regarding the potential roles, benefits, and limitations of LLMs in decision-making processes.

Future research can explore a larger and more diverse sample of participants, incorporate additional evaluation metrics, and investigate other decision-making domains to further enhance our understanding of the capabilities and limitations of LLMs [13]. This will facilitate the responsible integration of AI technologies in decision-making processes, ensuring their effective utilization while preserving human judgment and ethical considerations.

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: 34-47.
- Byerly S, Maurer L R, Mantero A, Naar L, An G, Kaafarani H M A (2021) Machine 2.Learning and Artificial Intelligence for Surgical Decision Making. Surgical Infections 22: 626-634.
- Garousi Mokhtarzadeh N, Amoozad Mahdiraji H, Jafari-Sadeghi V, Soltani A, Abbasi Kamardi, AA (2020) A producttechnology portfolio alignment approach for food industry: a multi-criteria decision making with z-numbers. British Food Journal 122: 3947-3967.
- 4. Grewal D, Roggeveen A L, Nordfält J (2017) The Future of Retailing. Journal of Retailing 93: 1-6.
- 5. Jarrahi M H (2018) Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Business Horizons 61: 577-586.
- Loftus T J, Tighe P J, FilibertoA C, Efron P A, Brakenridge S C, et al. (2020) Artificial Intelligence and Surgical Decisionmaking. JAMA Surgery 155: 148-158.
- Losservatore Romano (2020) Robotics at the service of the common good. L'Osservatore Romano https://www. osservatoreromano.va/it/news/2020-11/quo-257/la-roboticaal-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.
- 9. Massaro M, Dumay J, Bagnoli C (2017) When the investors speak: intellectual capital disclosure and the Web 2.0. Management Decision 55: 1888-1904.
- 10. Min H (2009) Artificial intelligence in supply chain management: theory and applications. International Journal of Logistics Research and Applications 13: 13-39.

- Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P (2021) The role of artificial intelligence in healthcare: a structured literature review. BMC Medical Informatics and Decision Making 21: 1-23.
- 12. Shrestha Y R, Ben Menahem S M, von Krogh G (2019) Organizational Decision-Making Structures in the Age of Artificial Intelligence. California Management Review 61: 66-83.
- 13. TalwarR, Koury A (2017) Artificial intelligence the next frontier in IT security? Network Security 2017: 14-17.

Copyright: ©2023 Mirza Niaz Zaman Elin. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.