Koordinat **Jurnal Pembelajaran Matematika dan Sains**
Volume 7, Number 1, January-June 2026, Pages 71-82

https://doi.org/10.24239/koordinat.v7i1.195

e-ISSN : 2745-4215

ANALYSIS OF CHATGPT AND GEMINI MENTAL MODELS OF APPARENT MOTION OF THE SUN AND STARS

Nurhikma Safitri[1*], Jusman Mansyur[1], Haeruddin[1] , Nurul Kami Sani[1], Andi Ulfah Khuzaimah[1]

[1] Department of Physics Education, Tadulako University, Central Sulawesi, Indonesia

ARTICLE INFO

Article History
Received: 20 Nov 2025
Revised: 29 May 2026
Accepted: 04 May 2026
Published: 09 Jun 2026

Keywords:
Apparent Motion of Stars
Apparent Motion of the Sun
ChatGPT
Gemini
Mental Model

ABSTRACT

This study aims to compare ChatGPT and Gemini's mental models regarding the apparent motion of the Sun and stars. The method used is a qualitative approach with an exploratory descriptive study using the reflective thematic analysis method according to Braun, Clarke, and Rance (2014). The subjects of this study were ChatGPT and Gemini, who also acted as respondents. The instrument used was 10 multiple-choice questions about the apparent motion of celestial bodies. Each answer was analyzed thematically to identify reasoning patterns, conceptual structures, and conformity with scientific mental models. The results show that ChatGPT predominantly displays a scientific mental model consistent with a heliocentric understanding and scientific explanation of apparent motion, while Gemini tends to display a hybrid mental model with a mixture of scientific concepts and preconceptions. The comparison of the two shows differences in conceptual depth, consistency of reasoning, and accuracy of scientific representation. These findings contribute to the development of the study of artificial intelligence mental models and open up opportunities for their use in AI-based science learning.

Corresponding Author:
Nurhikma Safitri, Department of Physics Education, Tadulako University, Central Sulawesi, Indonesia
Email: nurhikmasafitri17@gmail.com

How to cite:
Safitri, N., Mansyur, J., Haeruddin, Sani, N. K., Khuzaimah, A. U. (2025). Analysis of ChatGPT and Gemini Mental Models of Apparent Motion of the Sun and Stars. *Koordinat Jurnal MIPA, 7*(1), 71-82. https://doi.org/10.24239/koordinat.v7i1.195

INTRODUCTION

The development of artificial intelligence (AI) technology has given rise to various Generative AI-based systems, which are technologies capable of automatically generating text, images, and responses based on data patterns learned during the training process. Two examples of Generative AI that are currently developing rapidly are ChatGPT developed by OpenAI and Gemini developed by Google DeepMind. Both systems work using a Large Language Model (LLM) based on the Transformer architecture trained with vast amounts of text data, enabling them to understand context, predict words, and produce responses resembling human language. These capabilities have led to Generative AI being widely used in education, including as a virtual tutor, feedback provider, and supporter of science learning (Tom B. Brown et al, 2020).

Research related to the use of Chat Generative Pre-trained Transformer Artificial Intelligence (ChatGPT AI) in education improves student engagement and learning outcomes. ChatGPT can answer questions, provide additional explanations, and assist in problem solving, thereby improving the quality of learning. In addition, ChatGPT provides consistent learning support that is accessible at any time, enabling students to learn independently and reducing their dependence on teachers. This implementation also creates an inclusive learning environment that supports students who are reluctant to ask questions in class (Anderson et al, 2023).

Martinez (2022) evaluated the effectiveness of ChatGPT in helping creative writers develop ideas, construct plots, and improve writing style. The study showed a significant improvement in the quality of students' writing and creativity. By utilizing suggestions and feedback from ChatGPT, students can explore various writing styles and different narrative structures, thereby enriching their learning experience in the field of creative writing.

ChatGPT demonstrates how chatbots can be used as virtual tutors that provide personalized learning assistance. ChatGPT is able to understand and respond to student questions on various subjects, adjusting the level of difficulty and teaching methods based on individual needs, as well as providing flexibility in learning according to the student's schedule, supporting holistic academic development (Brown et al, 2023). In addition to ChatGPT, Gemini also has great potential in the world of education. With early detection, educators can design more effective learning strategies tailored to students' needs. For example, if Gemini finds that a student is having difficulty understanding a particular mathematical concept, the teacher can provide special attention or customized material to help that student. This early intervention is crucial to prevent learning difficulties from becoming more serious problems in the future (Wang et al, 2023).

Gemini is used to predict student academic performance by analyzing historical data and factors that influence learning outcomes. The predictive algorithm developed has higher accuracy than traditional methods, helping educators identify students who need help before a significant decline in performance occurs. These predictions enable schools to design proactive intervention programs and support curriculum development and teaching strategies (Gupta & Singh, 2023).

Khan & Yousuf (2019) examined the role of artificial intelligence (AI) in helping students understand complex astronomical concepts, such as planetary motion, lunar phases, and the apparent motion of celestial bodies. The results of this study show that AI is able to identify students' misconceptions related to these phenomena and provide relevant contextual corrections. Through AI-based interactions, students receive more detailed explanations and dynamic visualizations that facilitate the formation of more accurate mental models. In addition, AI can provide feedback tailored to each student's level of understanding, making it effective in correcting misconceptions and improving understanding of astronomy. Smith & Jones (2020) conducted research on mental models in science education, discussing how students understand the concept of apparent motion of the sun and

stars. They found that students' understanding is often influenced by inaccurate mental models. In their study, students often combined incorrect everyday understandings with the scientific information taught, thereby creating concepts that were not entirely accurate in understanding the apparent motion of celestial bodies. This study identified students' misconceptions but did not offer practical solutions to correct these misunderstandings.

Brown (2021) discusses the potential of AI, such as ChatGPT and Gemini, in helping students build more accurate mental models. Brown highlights that text-based interactions with AI can correct students' misconceptions about scientific concepts, including astronomy, in a more structured and adaptive way. It does not discuss the limitations of AI in providing in-depth explanations. Vosniadou & Brewer (1992) show that many students have incorrect mental models about the apparent motion of the sun and stars, often maintaining a geocentric view even after being given accurate scientific explanations. The limitations of visualizations in textbooks, which present abstract concepts statically, contribute to this problem. However, artificial intelligence technologies such as ChatGPT and Gemini can be utilized to provide interactive explanations tailored to the user's understanding. These AI systems have the potential to correct misconceptions through dynamic visualizations and adaptive responses, thereby helping to form more accurate mental models and enhance the effectiveness of science learning. Bekaert (2022) conducted research and discovered various mental models used by students to explain the phenomenon of apparent motion, in which many students mistakenly assume that the trajectory of stars is similar to that of the sun. This shows that although apparent motion is often seen in everyday life, understanding its scientific causes remains a challenge. This study emphasizes the need for more in-depth instruction to help students distinguish between the apparent motion of the sun and stars. In addition, analyzing artificial intelligence mental models such as ChatGPT and Gemini in understanding this astronomical concept can provide insight into whether AI experiences the same difficulties as humans or can offer a different approach.

Human mental models are formed through direct experience, sensory perception, and complex cognitive processes, whereas AI models such as ChatGPT and Gemini are based on algorithms and statistical data. This makes mental models in AI more training data-based and incapable of experiencing phenomena like humans do (Russell & Norvig, 2020).

Duit & Treagust (2003) mental models play an important role in science learning, as they help students build internal representations that enable them to understand and predict phenomena in the real world. In the context of the apparent motion of the sun and stars, this approach allows students to link observational experiences with scientific explanations, such as how the Earth's rotation affects the apparent position of the sun and stars in the sky. Vosniadou (1994) explains that understanding scientific concepts is often hindered by initial mental models formed by students based on intuition or everyday experiences. Misconceptions that arise, such as the assumption that the sun revolves around the Earth, can be corrected by providing students with opportunities to reconstruct their mental models through exploration-based learning and visualization. Both views emphasize the importance of using mental models to support a deep understanding of the concept of apparent motion in astronomy.

The apparent motion of the sun and stars is an important phenomenon in astronomy and celestial mechanics. Williams (2020) explains that the daily motion of the sun and stars in the sky can be explained by the relative motion of the Earth. Observational data shows that this movement pattern can be understood using the concept of parallax, which is an important tool for measuring the distance between stars.

Khurshid (2019) compares the apparent motion of the sun and identified stars, revealing similarities and differences

in movement patterns. The author uses mathematical models to explain the influence of gravity on this motion and introduces the Doppler effect phenomenon, which can be observed in moving stars.

Tran (2021) discusses apparent motion in the context of celestial mechanics, highlighting how the relative motion between the Earth, Sun, and stars causes this phenomenon. The computer simulations presented in this study provide a clear visualization of apparent motion, enhancing our understanding of the interactions between celestial objects. Patel (2018) highlights the impact of Earth's rotation on the apparent motion of stars, explaining that this rotation causes changes in the position of stars visible in the night sky. The author provides a physical explanation of angular velocity and how it affects astronomical observations. Fletcher (2022) examines the relationship between the motion of the sun and the motion of stars in the galaxy, using data from modern observatories. This research shows that apparent motion is not only important for understanding the movement of individual celestial objects but also for explaining the distribution of stars and the overall structure of the galaxy.

Based on the above description, it appears that research on ChatGPT still focuses on how ChatGPT improves engagement, learning outcomes, and personalization of education, as does Gemini. Research on mental models in general still focuses on students or university students as research subjects. There has been no research examining the combination of AI and mental models. Therefore, researchers are interested in conducting comparative research on the mental models of ChatGPT and Gemini regarding the apparent motion of the sun and stars.

METHOD

This research uses a qualitative approach with exploratory-descriptive study to identify and compare the mental models possessed by ChatGPT and Gemini in the context of the apparent motion of the sun and stars. The research design used in this study employs an exploratory-descriptive approach with qualitative methods. This study aims to identify and compare the mental models of two AI systems, namely ChatGPT and Gemini, in understanding the concept of the apparent motion of the sun and stars. A qualitative approach is used to analyze in depth how both AIs respond to questions related to this topic.

The research was conducted at the Physics Education Study Program, Tadulako University, in the even semester of the 2024/2025 academic year. The subjects of this study are ChatGPT and Gemini. The unique features possessed by AI models such as ChatGPT and Gemini make them interesting subjects for comparative analysis in the context of AI technology development (Rane et al, 2024). Each AI model, namely ChatGPT and Gemini, was tested using the same 10 questions. For each AI, each question was asked 25 times in separate conversations, resulting in 250 responses per model (25 repetitions × 10 questions). Thus, the analysis was conducted on two separate data sets, each consisting of 250 responses, representing the response patterns of ChatGPT and Gemini separately.

The instrument used in this study consisted of ten multiple-choice questions about the apparent motion of the sun and stars. The questions were submitted to the AI using standard prompts or commands that were carried out consistently in two stages, namely: (1) selecting answers, and (2) explaining the reasons. This instrument aimed to evaluate the extent to which each AI was able to provide consistent and accurate explanations.

Data collection in this study was carried out in three main stages to ensure that the data obtained was accurate, complete, and in accordance with the analysis requirements. The first stage was the preparation stage, which included searching for and collecting relevant literature, determining the location and subject of the study, compiling the instruments and prompts (instructions) to be used, and validating the instruments with two lecturers who acted as expert validators to ensure the suitability and quality of the instruments. The second stage was the implementation stage, which involved collecting data by asking conceptual questions to two artificial

intelligence models, namely ChatGPT and Gemini, according to the procedures that had been designed beforehand. The third stage was the final stage, which was the process of processing and analyzing the data that had been obtained, followed by presenting the research results in the form of a systematic and comprehensive final report.

This study uses testing techniques to examine the conceptual responses of two artificial intelligence models, ChatGPT-3.5 and Gemini 2.5, to ten multiple-choice questions designed to reveal their understanding of the apparent motion of the sun and stars. Each question was asked 25 times repeatedly to each model through separate conversation sessions to ensure the independence of responses. The question submission procedure was carried out in two stages: the first stage asked the AI to choose one of five answer options (a–e), and the second stage asked for a brief explanation as the basis for selecting the answer. Separate conversation sessions were used as a control so that each response represented the result of independent processing without the influence of previous interactions.

Data analysis in this study was conducted using a reflective thematic analysis approach as developed by Braun, Clarke, and Rance (2014). This approach was chosen because it provides a systematic yet flexible framework for exploring the meaning contained in qualitative data, as well as enabling the identification of conceptual thinking patterns from textual responses generated by artificial intelligence models.

The analysis process was carried out in stages through the following six main steps, first Familiarization with the Data. The researcher repeatedly read all the data from the ChatGPT-3.5 and Gemini 2.5 responses, both in the form of answer choices and reasons given, in order to gain a comprehensive understanding of the content and context of the meanings conveyed. Generating Initial Codes, Each unit of meaning in the AI's answers was systematically coded. These codes included important concepts, reasoning patterns, or key terms that reflected how the AI interpreted the apparent motion of the sun and stars. Searching for Themes, Similar codes are analyzed and grouped into initial themes. These themes represent the form or structure of the mental model adopted by each AI, such as the Earth-based rotation model, the geocentric model, or a hybrid representation. Reviewing Themes, the compiled themes are reviewed to ensure that each theme is strongly supported by the data and does not overlap with other themes. This process also involves checking whether the themes are representative of the entire dataset. Defining and Naming Themes Each theme is given a name and operational definition that reflects the key characteristics of the mental model found. The naming of these mental models refers to relevant literature and is adapted to the context of the findings in each AI. Producing the Report, the final findings are compiled in the form of an analytical narrative supported by representative quotations from AI responses. Each theme is associated with a mental model classification, namely initial model, hybrid model, and scientific model, to assess the level of conceptual depth of understanding of each AI. The classification of this study is divided into the following three categories; (1) Initial Model (InM): Both the choice of answer and the reasoning provided show misconceptions or inconsistencies with scientific concepts. (2) Hybrid Model (HyM): Only one element is correct, i.e., the answer is correct but the reasoning is wrong, or vice versa. (3) Scientific Model (ScM): Both the answer and the reasoning are consistent and in accordance with correct scientific understanding. Next, the percentage for each criterion is calculated as follows.

$$\text{Presentase} = \frac{Jumlah\ Skor}{Skor\ Maksimum}\ x\ 100\%$$

The analysis was conducted separately for each AI model so that the cognitive characteristics of ChatGPT and Gemini could be comprehensively identified without overlapping. The entire analysis process was reflective and interpretive, in which the researcher actively constructed meaning from the data based on context,

consistency, and logical connections between responses.

RESULT AND DISCUSSION

Comparison of ChatGPT and Gemini Mental Models

The following are the results of a comparison between ChatGPT and Gemini based on the distribution of mental models that emerged from the overall responses. The recapitulation shows that ChatGPT tends to produce scientific mental models more dominantly than Gemini, while Gemini more often displays initial model and hybrid model mental models. These findings indicate a significant conceptual difference between the two AIs in explaining the phenomenon of apparent motion of the sun and stars.

Comparison of ChatGPT and Gemini Mental Model Categories

The following bar chart presents the distribution of mental model classifications demonstrated by ChatGPT and Gemini based on 10 questions about the concept of apparent motion of the sun and stars. Each AI was given 25 opportunities to answer each question, resulting in a total of 250 responses from each model.

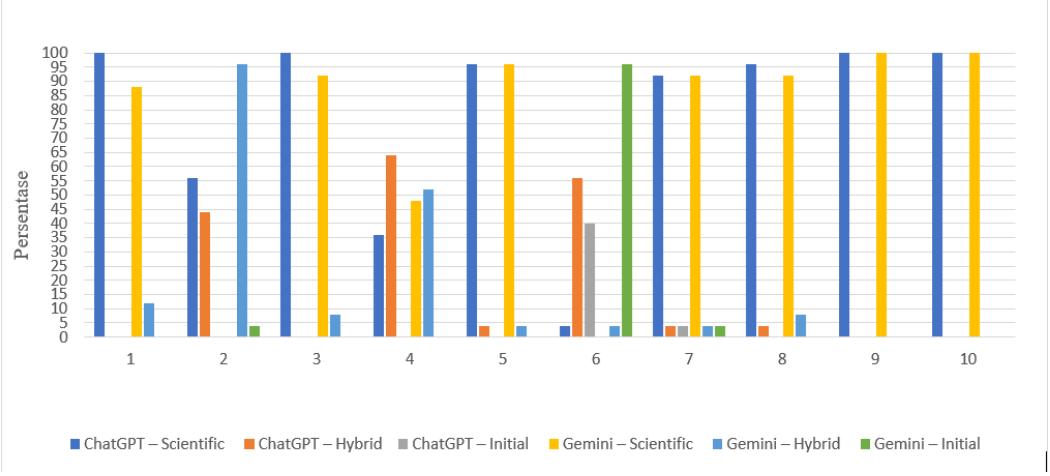


Figure 1. Percentage Comparison of ChatGPT and Gemini Response Values.

Based on Figure 1, which shows a bar chart comparing the distribution of ChatGPT and Gemini mental model categories based on ten assessment items. Mental model categories are divided into three: scientific, hybrid, and initial, each represented by a different color for both AIs. The graph shows that ChatGPT is more dominant in the scientific category in almost all questions, as indicated by the consistently high dark blue bars, for example in questions 1, 3, 7, 8, 9, and 10, which are close to or reach 100%. This shows that ChatGPT tends to provide answers that are in line with the scientific mental model and are consistent in most questions.

In contrast, Gemini is more prevalent in the initial model category, as seen from the very dominant yellow bar in questions 1 to 10. This tendency shows that Gemini often provides answers with a simple or less scientific initial mindset, so the distribution of the scientific category for Gemini is relatively low, except in certain questions. The hybrid category for both AIs is seen in much smaller numbers and scattered across several questions, such as questions 2, 4, and 6. This indicates the presence of answers that are a mixture of initial and scientific mindsets, but not dominant.

Overall, this graph shows a clear pattern where ChatGPT excels in generating scientific mental models, while Gemini tends to stick to initial mental models. This finding reinforces the per-question analysis results that ChatGPT is more consistent in understanding the concept of apparent motion of the sun and stars compared to Gemini.

Total Mental Model Recap

The distribution of mental model categories generated by ChatGPT and Gemini is presented to show the tendencies

of both in each question. A summary of the comparison of mental model categories and their percentages is shown in Figure 2.

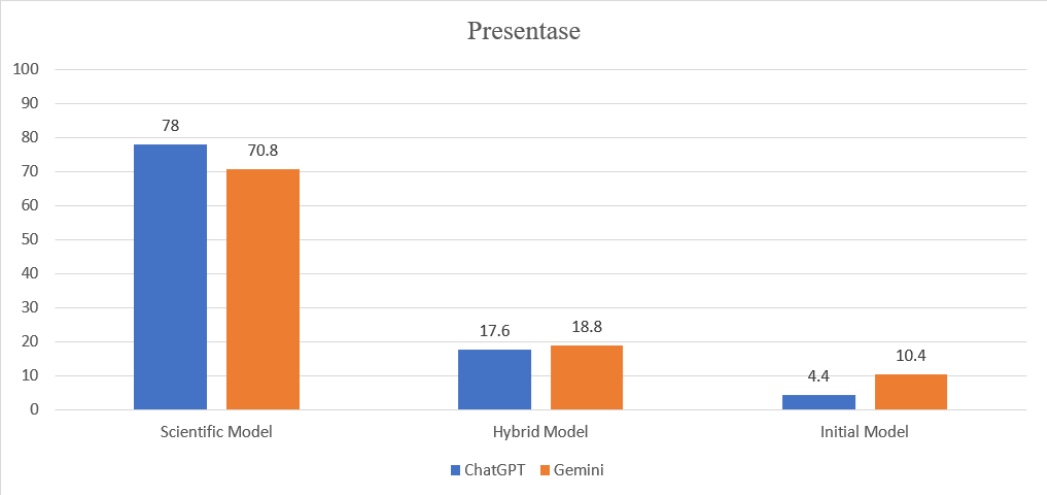


Figure 2. Distribution of Mental Model Categories

The distribution shows that ChatGPT is more dominant in generating scientific models (78%) than Gemini (70.8%). Although the percentage of hybrid models is relatively comparable (ChatGPT: 17.6% vs Gemini: 18.8%), a striking difference can be seen in the emergence of initial models, namely representations that contain misconceptions. Gemini shows a higher tendency to generate initial models (10.4%) than ChatGPT (4.4%). This indicates that Gemini tends to be more susceptible to reasoning that deviates from scientific concepts in some cases.

Comparative Analysis of AI Mental Models

Based on the total distribution of mental models from the two artificial intelligence models studied, namely ChatGPT and Gemini, there are differences in their abilities to understand astronomical concepts related to the apparent motion of the sun and stars.

ChatGPT consistently shows a dominance of scientific mental models with a percentage of 78%, higher than Gemini's 70.8%. This reflects that ChatGPT more often provides answers and reasons that are fully in line with scientific concepts. Additionally, ChatGPT also shows a relatively low initial model rate (4.4%), indicating a low tendency for this AI to provide responses that completely deviate from scientific principles. Conversely, although Gemini recorded a hybrid model proportion almost equivalent to ChatGPT (18.8% compared to 17.6%), this AI produced an initial model of 10.4%, more than double that of ChatGPT. This shows that Gemini more often displays reasoning patterns that contain misconceptions or fundamental inconsistencies with astronomical concepts.

This difference in performance becomes even clearer when analyzed at the individual question level. Gemini shows the most striking weakness by generating 96% initial models, indicating that most of its responses contain fundamental misconceptions. In contrast, ChatGPT, while not entirely correct, still displays significant scientific and hybrid models. On the other hand, for questions that require temporal logic or an understanding of the Earth's rotation and the changing positions of celestial bodies, both AIs were able to provide completely scientific answers, showing that they both have strong competence in answering questions based on temporal causality or daily motion.

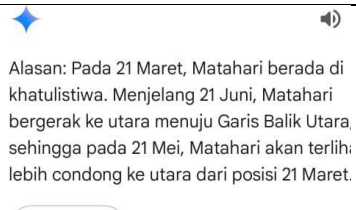
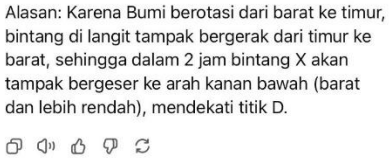
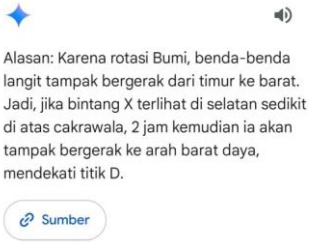
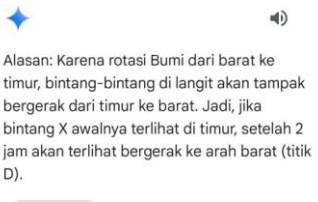
Thus, it can be concluded that ChatGPT tends to be more stable in

maintaining scientific reasoning across various types of questions, while Gemini shows more fluctuating performance. Gemini can provide very accurate answers in some situations, but it also has the potential to generate completely inappropriate responses to certain questions that require complex spatial and conceptual integration.

Comparative Analysis of Mental Model Categories

The following comparative analysis of mental model categories aims to examine the conceptual tendencies of ChatGPT and Gemini in greater depth. The results of the analysis show that ChatGPT is more consistent in producing scientific mental models, while Gemini tends to vary with a greater proportion in the initial model and hybrid model categories. These findings confirm that there are clear conceptual differences between the two AIs in representing scientific phenomena.

Table 1. Comparative Analysis of Mental Model Categories

Mental Model Category	ChatGPT	Gemini	Analysis
Scientific			Demonstrates the concept of Earth's rotation accurately.
Hybrid			The concept is correct, but it's mixed with misconceptions.
Initial			The explanation contradicts astronomical concepts.

From the above analysis, it is evident that ChatGPT excels in generating comprehensive and stable scientific models, particularly due to its ability to consistently mention and integrate technical astronomical terms such as declination, equinox, solstice, apparent daily motion, and observer latitude. Answers categorized as scientific models by ChatGPT are almost always accompanied by causal arguments that are consistent with astronomical principles. In contrast, Gemini tends to be more concise and descriptive, and experiences more inconsistencies in explaining the direction or dynamics of celestial motion. This is evident in the hybrid model and initial model categories, where Gemini more often mixes scientific knowledge with everyday intuition, or demonstrates errors in distinguishing between the concepts of rotation and revolution.

However, it should be noted that both AIs perform very well on some questions, with scientific reasoning levels approaching perfection. This indicates that both ChatGPT

and Gemini have scientific representation potential, but ChatGPT's conceptual structure stability is more consistent than Gemini's.

Discussion

This study identifies three types of mental models formed from AI responses to conceptual astronomy questions, namely: initial models, hybrid models, and scientific models. This classification refers to the general framework in mental model studies, as developed by Vosniadou & Brewer (1992) and Treagust et al. (2002), and reinforced through reflective thematic analysis according to the approach of Braun, Clarke, and Rance (2014).

The initial model is an early conceptual representation that is completely inconsistent with scientific understanding. In this study, Gemini showed a higher tendency to produce initial models (10.4%) than ChatGPT (4.4%). For example, in understanding the spatial motion of stars in the night sky, Gemini produced (96%) responses categorized as initial models, reflecting a fundamental misconception. The finding regarding the dominance of initial models in Gemini's responses is the main focus of this study's analysis.

These findings are consistent with the results of Hans Bekaert et al. (2022) and Plummer (2009), which show that even students who have taken astronomy courses still often use incorrect representations when explaining the apparent motion of celestial bodies. Similarly, Vosniadou & Brewer (1992) found that children tend to form naive models, such as the belief that the sun hides behind mountains, which is an early form of the initial model. Gemini's tendency to produce similar answers to questions that require spatial-conceptual integration shows that artificial intelligence can also reflect unscientific human thought structures.

The hybrid model (previously called the synthetic model) is a category that represents a mixture of scientific understanding and misconceptions. Both ChatGPT (17.6%) and Gemini (18.8%) showed the emergence of this model in almost equal proportions. In questions 2 and 4, for example, both AIs tended to give correct answers with incorrect reasoning, or vice versa.

The presence of this mixed mental model reinforces the findings of Vosniadou & Skopeliti (2014), which explain that the transition from misconceptions to scientific understanding does not occur suddenly, but rather through a synthetic phase. In humans, this model often arises due to conflicts between intuitive knowledge and new scientific information. In the context of AI, the emergence of hybrid models likely reflects the integration of contradictory data in the training corpus or an inability to filter consistent logic when generating reasoning.

Furthermore, these results expand on the findings of Schulze Balhorn et al. (2024), who discovered that although many of ChatGPT's answers were rated as mostly correct, the reasoning behind these answers was not always critical and could contain weaknesses in reasoning. A similar pattern was observed in Gemini's answers, which were sometimes correct in terms of choice but weak in justification.

A scientific model is a form of conceptual representation that aligns with scientific concepts and principles. In this study, ChatGPT produced scientific models 78% of the time, while Gemini did so 70.8% of the time. Scientific models were particularly dominant in questions that required the logic of the daily motion of celestial bodies (e.g., questions 1, 9, and 10), whose answer structures were relatively linear and did not require complex spatial transformations.

The dominance of scientific models in ChatGPT shows that this system is capable of constructing complete and coherent scientific reasoning. This supports the findings of Brown & Thompson (2022), which show that the use of generative AI can help strengthen understanding of physics and science concepts when used appropriately. In this context, ChatGPT shows potential as a reliable learning assistant in explaining basic astronomy concepts.

However, it should be noted that these results do not fully guarantee ChatGPT's reliability in all contexts. A study by ALZu'bi et al. (2023) confirms that although

ChatGPT can generate relevant text, this model still faces challenges in understanding certain contexts and avoiding bias. This means that the presence of a scientific model is not only a reflection of AI's capacity, but also a product of how this system is trained and commanded.

In general, these findings support the results of Nyaaba's (2023) study, which shows that ChatGPT and Gemini can provide different responses to scientific concepts, with ChatGPT tending to be concise and Gemini more complex but not always accurate. This difference is evident in the distribution of mental models found in this study. ChatGPT is more stable in producing scientific models, while Gemini is more prone to initial models and hybrid models.

Meanwhile, Smith & Johnson (2023) highlight Gemini's strengths in terms of interactivity and real-time data access, but this study shows that in the context of conceptual reasoning about astronomy, Gemini is not yet fully superior. Weaknesses in basic reasoning structures are still evident in the form of misconceptions that appear in initial models.

This study has several limitations that need to be considered in the development of similar studies in the future. First, the research objects only covered two artificial intelligence models, namely ChatGPT and Gemini, so the analysis results cannot be generalized to other artificial intelligence models that have different characteristics and knowledge systems. Second, the research instruments were limited to ten questions used to explore the mental models of the two AI, so the conceptual variations produced were still limited. Third, the analysis method used is qualitative, referring to a specific assessment rubric, so that data interpretation is highly dependent on the accuracy and objectivity of the researcher in assessing the clarity, accuracy, understanding, and consistency of the concepts presented by each AI model.

CONCLUSION

Based on the results of the research conducted, it can be concluded that ChatGPT performs better in answering conceptual questions about the apparent motion of the sun and stars compared to Gemini. This can be seen from ChatGPT's tendency to more frequently produce scientific mental models with clear, accurate, and consistent explanations. ChatGPT is able to establish conceptual relationships that are consistent with the principles of Earth's rotation and revolution and their implications for astronomical phenomena.

Gemini tends to produce hybrid or initial mental models, which are characterized by less accurate explanations that are not always consistent and still contain a mixture of scientific and intuitive understanding. Thus, ChatGPT is superior in representing scientific mental models that are in line with actual astronomical concepts, while Gemini shows a more varied but less stable understanding.

ACKNOWLEDGMENT

With deep gratitude, the author would like to thank the various parties who have supported the author's journey. The author expresses sincere gratitude to his parents who have educated, guided, prayed, sacrificed, advised, and motivated him with patience and sincerity. Although the author realizes that he cannot fully repay them. Thanks also to my siblings who provided prayers, assistance, and motivation. Furthermore, my deepest gratitude goes to my advisor and lecturer who took the time to guide me and provide advice and direction that was very helpful. Thank you also to the first and second discussants who provided input and suggestions for the completion of this thesis. Not to forget, thank you to myself for fighting and never giving up in the face of every challenge. May all the kindness given be rewarded manifold and bring benefits in the future.

AUTHOR CONTRIBUTIONS

The First Author was responsible for conceptualization, writing the original draft, editing, and visualization. The Second Author acted as an advisor, providing direction, writing the literature review, methodological analysis, as well as validation and supervision. The Third,

Fourth, and Fifth Authors acted as advisors and supervised the entire research process. Each author made a significant contribution to the success of this research.

REFERENCES

Anderson, J., Smith, L., & Doe, R. (2023). Improving Student Engagement and Learning Outcomes with AI-Powered Chatbots. Journal of Educational Technology, 15(2), 45-67.

AlZu'bi, S., Mughaid, A., Quiam, F., & Hendawi, S. (2023). Exploring the Capabilities and Limitations of ChatGPT and Alternative Big Language Models. Artificial Intelligence and Applications. DOI: 10.47852/bonviewAIA3202820

Brown, T., Harris, M., & Walker, K. (2023). AI-Powered Tutoring Systems: The Role of ChatGPT in Personalized Education. Journal of Innovative Education, 11(2), 150-170.

Braun, Virginia, Victoria Clarke, and Nicola Rance. 2014. How to use thematic analysis with790 interview data. In The counselling and psychotherapy research handbook, eds.791 Andreas Vossler, and Naomi Moller, 183-197. London, UK: Sage Publications.

Brown, A. (2021). The role of AI in science education: Building mental models through technology. Artificial Intelligence in Education, 48(2), 123-134.

Bekaert, H., Van Winckel, H., Van Dooren, W., Steegen, A., & De Cock, M. (2022). Identifying students' mental models of the apparent motion of the Sun and stars. Physical Review Physics Education Research, 18(1), 010130.

Brown, J., & Thompson, R. (2022). Enhancing problem-solving skills through generative AI in educational settings. Journal of Educational Technology, 20(4), 45-60.

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*(6), 671–688.

Fletcher, R. J. (2022). Analysis of Solar Motion in Relation to Stellar Motion. Astronomy and Astrophysics Review, 30(4), 315-330.

Gupta, R., & Singh, A. (2023). Predicting Academic Performance with Gemini: A Machine Learning Approach. Educational Data Science Journal, 9(4), 210-225.

Khan, M. A., & Yousuf, M. (2019). Exploring the Role of Artificial Intelligence in Science Education: A Case Study in Astronomy Learning. Journal of Educational Technology Systems, 48(2), 214-227.

Khurshid, M. A. (2019). Solar and Stellar Motion: A Comparative Study. *Journal of Modern Physics*, 10(5), 500-510.

Martinez, A., Gonzalez, P., & Lee, S. (2022). Enhancing Creative Writing through AI: A Study on the Impact of ChatGPT on Writing Quality. International Journal of Creative Writing, 8(3), 120-135.

Nyaaba, M. (2023). Comparing Human and AI's (GPT-4 and Gemini) Understanding of the Nature of Science. SSRN Electronic Journal.

Patel, K. R. (2018). The Role of the Earth's Rotation in the Apparent Motion of Stars. International Journal of Astronomy, 6(3), 45-55.

Plummer, J. D. (2009). Children's understanding of the apparent motion of celestial bodies. Physical Review Special Topics - Physics Education Research, 5(2), 020103.

Wang, X., Zhao, L., & Chen, H. (2023). Leveraging Gemini AI for Early Diagnosis of Learning Disabilities. Journal of Learning Disabilities Research, 12(1), 78-92.

Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Boston: Pearson. ISBN: 978-0134610993.

Smith, D., & Jones, R. (2020). Student misconceptions in understanding apparent motion of celestial bodies.

International Journal of Science Education, 42(1), 47-63.

Smith, J., & Johnson, A. (2023). Comparative Analysis of AI Models: ChatGPT, Gemini, and Perplexity AI. Journal of Artificial Intelligence Research, 45(2), 123-145.

Schulze Balhorn, L., Weber, J. M., Buijsman, S., Hildebrandt, J. R., Ziefle, M., & Schweidtmann, A. M. (2024). Empirical assessment of ChatGPT's answering capabilities in natural science and engineering. Scientific Reports, 14(1), 4998.

Tran, L. H. (2021). Understanding Apparent Motion in Celestial Mechanics. Celestial Mechanics and Dynamical Astronomy, 139(1), 87-102.

Vosniadou, S., & Brewer, W. F. (1992). Mental Models of the Earth: A Study of Conceptual Change in Childhood. In The development of scientific reasoning (pp. 81-94). Lawrence Erlbaum Associates.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. Learning and Instruction, 4(1), 45–69.

Williams, J. D. S. (2020). The Apparent Motion of the Sun and Stars. Astrophysics Journal, 150(2), 123-130.