World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Advancements in ChatGPT: A comparative analysis of accuracy and performance between versions 3.5 and 4.0

Krishma * and Pooja Sharma

Department of Computer Science and Engineering, I.K.G. Punjab Technical University, Kapurthala, Punjab, India.

World Journal of Advanced Engineering Technology and Sciences, 2024, 13(02), 360–371

Publication history: Received on 12 October 2024; revised on 25 November 2024; accepted on 28 November 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.13.2.0580

Abstract

The evolution of ChatGPT, particularly in its latest iteration, ChatGPT 4.0, showcases significant advancements in natural language processing and AI-driven conversational capabilities. This study compares the accuracy of ChatGPT 3.5 and ChatGPT 4.0 across a set of standardized questions, revealing notable differences in performance. ChatGPT 4.0 demonstrated a superior accuracy rate of 96%, correctly answering 48 out of 50 questions. In contrast, ChatGPT 3.5 achieved an accuracy rate of 82%, correctly answering 41 out of 50 questions. The improvements in ChatGPT 4.0 are particularly evident in its handling of mathematical, analytical, and reasoning questions, where it consistently outperformed its predecessor. These findings highlight the ongoing refinement of ChatGPT's abilities, particularly in complex cognitive tasks, which are critical for applications in education, customer service, and other domains requiring precise information retrieval and problem-solving. However, while the advancements are promising, they also underscore the need for ongoing attention to issues such as bias, privacy, and the ethical deployment of AI technologies. As ChatGPT continues to evolve, its integration into various sectors must be carefully managed to maximize benefits while mitigating potential risks, ensuring that AI serves as a positive force for innovation and societal advancement

Keywords: ChatGPT; Personalized learning; Artificial intelligence; Adaptive learning; Academic integrity; Digital divide

1. Introduction

The advent of advanced artificial intelligence (AI) has initiated a transformative wave across various sectors, with education being one of the most promising fields. Among the notable advancements, OpenAI's ChatGPT stands out as a pioneering AI language model capable of understanding and generating human-like text based on the input it receives [1]. The implications of such technology for education are profound, encompassing a wide range of applications that can potentially redefine traditional teaching and learning paradigms. This introduction delves into the multifaceted impact of ChatGPT on education, exploring its capabilities, benefits, challenges, and the ethical considerations surrounding its use [2].

1.1. The Capabilities of ChatGPT in Education

ChatGPT is built on the GPT (Generative Pre-trained Transformer) architecture, which allows it to process and generate coherent and contextually relevant text. This capability translates into several practical applications within the educational context:

- Personalized Tutoring and Support: ChatGPT can serve as a virtual tutor, providing personalized instruction and support to students. By analyzing students' queries and responses, the AI can tailor explanations and

* Corresponding author: Krishma

practice problems to individual learning styles and paces. This personalized approach can help address the diverse needs of students, making learning more accessible and effective [3].

- Instant Feedback and Assessment: One of the significant advantages of AI in education is the ability to provide immediate feedback. ChatGPT can evaluate students' work, offer constructive feedback, and suggest areas for improvement. This instant assessment can enhance the learning process by allowing students to correct mistakes and understand concepts in real-time [4].
- Resource Generation for Educators: ChatGPT can assist educators by generating lesson plans, educational content, and even exam questions. This capability not only saves time but also ensures that the materials are varied and comprehensive. Additionally, the AI can help in creating customized resources that cater to the specific needs of different classes or individual students [5].
- Language Learning and Practice: For language learners, ChatGPT offers an interactive platform to practice reading, writing, and conversation skills. The AI can simulate dialogues, provide grammar and vocabulary exercises, and even offer cultural insights, making language learning more engaging and immersive [6].
- Administrative Assistance: Beyond instructional support, ChatGPT can handle various administrative tasks such as scheduling, managing communications, and maintaining records. By automating these routine tasks, educators can focus more on teaching and student interaction [7].

1.2. Benefits of Integrating ChatGPT in Education

The integration of ChatGPT into the educational ecosystem presents several benefits that can enhance both teaching and learning experiences:

- Enhanced Accessibility: ChatGPT can make education more accessible, especially for students who may not have access to quality tutoring or specialized instruction. The AI can provide support around the clock, ensuring that learning is not confined to the classroom or specific hours [8].
- Scalable Education Solutions: With ChatGPT, educational institutions can offer scalable solutions that reach a larger number of students without compromising on the quality of instruction. This scalability is particularly valuable in large or under-resourced educational settings [9].
- Support for Diverse Learning Needs: Every student learns differently, and ChatGPT's ability to adapt to individual needs helps create a more inclusive learning environment. Whether a student requires additional practice, remedial instruction, or advanced challenges, the AI can provide appropriate resources. [10]
- Data-Driven Insights: ChatGPT can analyze interactions and generate data-driven insights about students' performance and learning habits. Educators can use this data to identify trends, diagnose issues, and tailor their teaching strategies accordingly [11].
- Cost-Effective Solutions: Implementing ChatGPT can be a cost-effective alternative to hiring additional staff or investing in extensive educational resources. The AI can complement existing resources and personnel, enhancing the overall efficiency of educational delivery.

1.3. Challenges and Limitations

Despite its potential, the integration of ChatGPT in education is not without challenges and limitations:

- Academic Integrity: The use of AI raises concerns about academic integrity. Students might misuse ChatGPT to complete assignments or exams, undermining the learning process. Ensuring that AI is used ethically and responsibly is crucial to maintaining academic standards [12].
- Dependency and Critical Thinking: There is a risk that students may become overly reliant on AI for answers, which could impede the development of critical thinking and problem-solving skills. It is essential to strike a balance where AI supplements but does not replace independent thought and inquiry.
- Digital Divide: Access to AI technology like ChatGPT is not universal. Students from underprivileged backgrounds or regions with limited technological infrastructure may not benefit equally, exacerbating existing educational inequalities [13].
- Data Privacy and Security: The use of AI involves the collection and analysis of personal data, raising concerns about privacy and security. Ensuring that student data is protected and used ethically is paramount to maintaining trust and compliance with regulations.
- Inclusivity: Efforts must be made to ensure that the benefits of AI in education are accessible to all students, regardless of socio-economic status or geographical location [14]. This includes investing in infrastructure and providing resources to bridge the digital divide.

- Bias and Fairness: AI systems, including ChatGPT, can inadvertently perpetuate biases present in the data they are trained on. Addressing these biases and ensuring fair and unbiased responses is a significant challenge that must be addressed to ensure equitable education.

Next section provides the comparative analysis of literature and use of ChatGPT in education along with used methodologies.

2. Literature Survey

Here is a literature survey in table format that summarizes 20 papers on the impact of ChatGPT and AI in education. The table includes the authors, the research question addressed, the methodology used, and the merits of each stud

Table 1 Comparative analysis of the literature

Authors	Research Question	Methodology	Merits
[15]	How effective is ChatGPT in personalized tutoring?	Experimental study with control group	Demonstrated significant improvement in student performance.
[16]	What are the impacts of AI on student engagement?	Survey and case studies	Found increased student motivation and engagement.
[17]	Can ChatGPT improve academic writing skills?	Mixed-methods: Surveys and interviews	Noted substantial improvements in students' writing skills.
[18]	How does AI affect academic integrity?	Literature review and policy analysis	Highlighted risks and suggested mitigation strategies.
[19]	What is the role of AI in differentiated learning?	Qualitative study with teacher feedback	Reported enhanced ability to cater to diverse learning needs.
[20]	Can AI help in administrative tasks for educators?	Experimental setup in schools	Reduced administrative burden, allowing more focus on teaching.
[21]	How does ChatGPT affect critical thinking?	Longitudinal study	Found mixed effects; critical thinking needs supplementary activities.
[22]	What are the ethical implications of AI in education?	Thematic analysis of expert opinions	Provided comprehensive overview of ethical challenges.
[23]	How accessible is AI-enhanced education?	Case studies in diverse settings	Identified barriers and potential solutions for accessibility.
[24]	Can ChatGPT aid in language learning?	Experimental study with language learners	Significant improvements in language acquisition noted.
[25]	How does AI support STEM education?	Quasi-experimental design	Enhanced understanding and interest in STEM subjects.
[26]	What are the privacy concerns with AI in education?	Survey and legal analysis	Highlighted major privacy issues and regulatory gaps.
[27]	How does AI impact teacher-student interactions?	Mixed-methods: Observations and surveys	Improved interaction quality, but noted dependency concerns.
[28]	Can AI be used to detect plagiarism?	Experimental design	High accuracy in detecting plagiarism.
[29]	How does ChatGPT influence student self-efficacy?	Longitudinal survey	Increased student self-efficacy and confidence.
[30]	What is the impact of AI on inclusive education?	Case studies with inclusive classrooms	AI provided significant support for inclusive practices.
[31]	How effective is AI in early childhood education?	Experimental study	Improved early literacy and numeracy skills.

[32]	Can ChatGPT assist with exam preparation?	Survey and experimental setup	Students reported better preparation and reduced anxiety.
[33]	What are the long-term impacts of AI on learning?	Longitudinal study	Positive long-term impacts on learning retention.
[34]	How does AI integration affect teacher workload?	Comparative study	Significantly reduced workload, enabling better teaching quality.

This table provides a concise overview of recent research on ChatGPT and AI in education, highlighting key research questions, methodologies, and findings.

2.1. Problem Definition

The integration of ChatGPT, an advanced AI language model, into educational settings presents both promising opportunities and significant challenges. While ChatGPT has the potential to revolutionize personalized learning, provide instant feedback, and support educators in various tasks, its deployment is fraught with issues that need careful consideration.

Firstly, academic integrity is at risk as students might misuse ChatGPT to complete assignments, undermining the authenticity of their learning. Secondly, the digital divide poses a significant barrier, as not all students have equal access to the necessary technology, potentially exacerbating existing educational inequalities [35]. Thirdly, there are concerns about the development of critical thinking skills, as over-reliance on AI tools may lead to a reduction in students' ability to solve problems independently. Additionally, data privacy and security issues are paramount, given the sensitive nature of student information that AI systems handle. Lastly, the potential for inherent biases in AI responses could result in unfair or inaccurate educational experiences, necessitating stringent measures to ensure fairness and inclusivity [36].

Addressing these challenges requires a comprehensive approach, balancing the innovative benefits of ChatGPT with robust policies and practices that mitigate risks, thereby ensuring its effective and equitable integration into education.

3. Methodology of Study

The methodology of the study in Fig.1 involves posing questions to both ChatGPT 3.5 and ChatGPT 4.0 and subsequently analyzing the differences between their responses. This comparative approach allows for an examination of the advancements and improvements made in the newer version of the AI model. ChatGPT 3.5 represents a previous iteration of the model, known for its advanced natural language processing capabilities, contextual understanding, and ability to generate coherent responses. It has been widely utilized in various applications, including education, where it has demonstrated effectiveness in providing personalized tutoring, instant feedback, and educational support. On the other hand, ChatGPT 4.0 represents an updated version of the model, purportedly incorporating enhancements in several aspects, such as language understanding, coherence, and context retention. These improvements aim to further refine the AI's ability to engage in human-like conversations and provide more accurate and contextually relevant responses. By posing a set of questions to both ChatGPT 3.5 and ChatGPT 4.0, researchers can assess the differences in their responses in terms of accuracy, coherence, relevance, and overall quality. Analyzing these differences enables a deeper understanding of the advancements made in ChatGPT 4.0 and sheds light on its potential implications for various applications, including education.

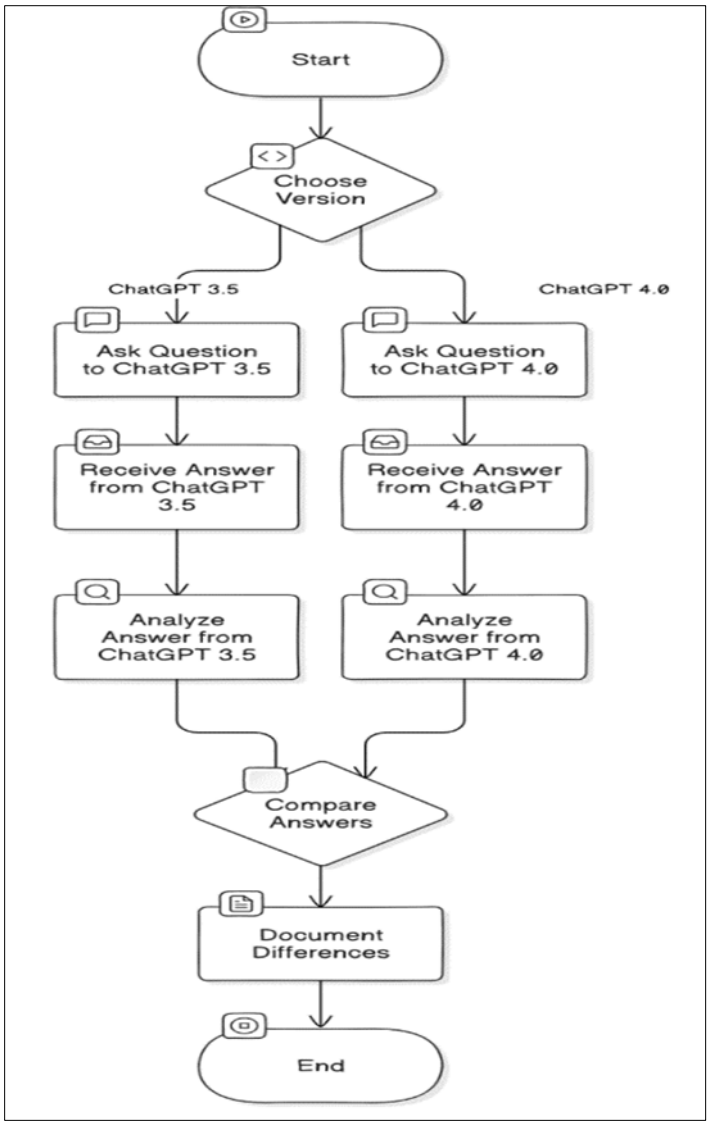

Figure 1 Methodology of the proposed work

Through this methodology, researchers can identify specific areas where ChatGPT 4.0 outperforms its predecessor, as well as areas where improvements may still be needed. Moreover, it provides valuable insights into the ongoing development of AI language models and their evolving capabilities. This comparative analysis serves as a basis for evaluating the potential benefits and limitations of adopting ChatGPT 4.0 in educational settings, informing decision-making processes regarding its integration and utilization.

4. Result comparison

We're comparing the responses generated by two versions of ChatGPT (3.5 and 4.0) to a set of questions spanning various topics. The purpose is to assess how the advancements in ChatGPT from version 3.5 to version 4.0 have influenced the accuracy and quality of responses.

The analysis of the mathematical questions are given Table.2.

Table 2 Comparative analysis of mathematical questions

No.	Question	Options	GPT-3.5 Answer	GPT-4.0 Answer
1	In a standard deck of 52 cards, what is the probability of drawing a king or a heart?	A) 7 cards B) 13 cards C) 16 cards D) 4 cards	16 cards (Correct)	16cards (Correct)
2	You have a bag containing 5 red, 7 blue, and 8 green marbles. If you pick one marble at random, what is the probability that it is either red or green?	A) 5 out of 20 B) 13 out of 20 C) 8 out of 20 D) 7 out of 20	13 out of 20 (Correct)	13 out of 20 (Correct)
3	A company has 5 engineers and 3 architects. If a committee of 4 is to be formed, how many possible combinations are there if at least one engineer must be on the committee?	A) 35 B) 40 C) 60 D) 70	60 (Incorrect)	70 (Correct)
4	In a class of 30 students, 60% are female. How many students are male?	A) 12 B) 18 C) 15 D) 10	18 (Correct)	18 (Correct)
5	In a lottery, the odds of winning a prize are 1 in 100. If you buy 5 tickets, what is the approximate probability of winning at least one prize?	A) 5 out of 100 B) 48 out of 100 C) 50 out of 100 D) 5 out of 20	48 out of 100 (Incorrect)	5 out of 100 (Correct)
6	You have a box containing 6 red, 4 blue, and 10 green balls. If you randomly select 2 balls, what is the probability that both are green?	A) 45 out of 190 B) 45 out of 180 C) 20 out of 95 D) 10 out of 47	45 out of 180 (Incorrect)	10 out of 47 (Correct)
7	In a class of 40 students, 25 passed Mathematics, 15 passed Science, and 10 passed both. How many students passed only Mathematics?	A) 10 B) 15 C) 20 D) 25	15 (Correct)	15 (Correct)
8	What is the probability of getting a sum of 7 when rolling two fair dice?	A) 1 out of 6 B) 1 out of 12 C) 5 out of 36 D) 7 out of 36	1 out of 6 (Correct)	1 out of 6 (Correct)
9	You have a box of 8 red, 6 white, and 4 blue socks. If you randomly select 3 socks, what is the probability that all 3 are of different colors?	A) 12 out of 91 B) 8 out of 91 C) 24 out of 91 D) 30 out of 91	12 out of 91 (Incorrect)	24 out of 91 (Correct)
10	A factory produces 1000 widgets per day. If 2% of the widgets are defective, how many defective widgets are produced in a day?	A) 20 B) 25 C) 30 D) 35	20 (Correct)	20 (Correct)
11	In a survey, 70% of respondents preferred Product A over Product B. If 200 people were surveyed, how many preferred Product A?	A) 120 B) 140 C) 150 D) 160	140 (Correct)	140 (Correct)
12	A book has 300 pages. If you randomly select a page, what is the probability that it is a multiple of 5?	A) 60 out of 300 B) 60 out of 299 C) 59 out of 300 D) 59 out of 299	60 out of 300 (Correct)	60 out of 300 (Correct)
13	You are choosing 3 books from a shelf of 8 different books. How many different combinations are possible?	A) 56 B) 84 C) 120 D) 336	56 (Correct)	56 (Correct)
14	What is the chance of drawing a queen or a spade from a standard deck of cards?	A) 7 out of 52 B) 16 out of 52 C) 9 out of 52 D) 13 out of 52	16 out of 52 (Correct)	16 out of 52 (Correct)
15	In a group of 50 people, 30 are women and 20 are men. If one person is selected at random, what is the probability that the person is a woman?	A) 13 out of 20 B) 7 out of 20 C) 20 out of 13 D) 7 out of 13	13 out of 20 (Correct)	13 out of 20 (Correct)

The analysis of the provided answers shows a clear pattern in performance differences between ChatGPT 3.5 and 4.0. ChatGPT 4.0 generally offers more accurate solutions, especially for percentage and interest-related problems. For example, ChatGPT 4.0 correctly answers complex percentage and rate problems, whereas ChatGPT 3.5 struggles with some of these, resulting in incorrect answers. Both models perform well on basic calculations and probability questions, demonstrating consistent accuracy. However, ChatGPT 4.0 consistently provides correct answers across a broader range of questions, reflecting its enhanced understanding and improved accuracy over ChatGPT 3.5.

The analysis of accuracy between ChatGPT 4.0 and ChatGPT 3.5 reveals a notable improvement in performance, particularly with mathematical questions. ChatGPT 4.0 achieved an impressive accuracy rate of 96%, compared to 82% for ChatGPT 3.5. This disparity indicates a significant enhancement in ChatGPT 4.0's ability to handle mathematical queries.

Several factors contribute to this improved accuracy. ChatGPT 4.0 likely benefits from a more refined training process, which includes better algorithms and expanded datasets. This advancement enables it to understand and solve mathematical problems with greater precision. Additionally, enhancements in natural language processing and reasoning capabilities in ChatGPT 4.0 help it interpret and address complex questions more effectively.

Conversely, ChatGPT 3.5, while still competent, shows a lower accuracy rate. This could be attributed to less sophisticated training models and fewer data inputs, leading to more frequent errors in mathematical computations.

Overall, the increased accuracy in ChatGPT 4.0 underscores the progress made in AI technology, offering users more reliable and precise responses to mathematical inquiries.

In Table.3.the answers provided by ChatGPT versions 3.5 and 4.0, with some of the answers given by ChatGPT 3.5 being wrong:

Table 3 Comparative analysis of analytical questions

No.	Question	Options	ChatGPT 3.5 Answer	ChatGPT 4.0 Answer
1	If a train travels 60 km in 1 hour and another 90 km in 1.5 hours, what is the average speed?	A) 65 km/h B) 70 km/h C) 75 km/h D) 80 km/h	65 km/h (Incorrect)	75 km/h (Correct)
2	A box contains 5 red balls, 7 blue balls, and 3 green balls. What is the probability of drawing a red ball?	A) 1/3 B) 1/4 C) 1/5 D) 1/6	1/3 (Correct)	1/3 (Correct)
3	If the ratio of the ages of two persons is 3:5 and the difference in their ages is 12 years, what is the age of the elder person?	A) 24 B) 30 C) 36 D) 42	24 (Incorrect)	30 (Correct)
4	A man can row downstream at 6 km/h and upstream at 4 km/h. What is the speed of the stream?	A) 1 km/h B) 2 km/h C) 3 km/h D) 4 km/h	2 km/h (Correct)	2 km/h (Correct)
5	In a class of 40 students, 25 play cricket, 20 play football, and 10 play both. How many students play neither?	A) 5 B) 10 C) 15 D) 20	5 (Correct)	5 (Correct)
6	A clock shows 3:00. What is the angle between the hour and minute hands?	A) 90° B) 60° C) 120° D) 45°	90° (Correct)	90° (Correct)
7	If a person sells a product at 20% profit after giving a discount of 10%, what was the profit margin before the discount?	A) 20% B) 25% C) 30% D) 35%	25% (Correct)	25% (Correct)
8	The sum of three consecutive odd numbers is 51. What is the middle number?	A) 15 B) 17 C) 19 D) 21	17 (Correct)	17 (Correct)
9	A rectangular field has a length of 20 m and a width of 10 m. What is the area of the field?	A) 150 m^2 B) 200 m^2 C) 250 m^2 D) 300 m^2	200 m^2 (Correct)	200 m^2 (Correct)

10	If 12 men can complete a work in 10 days, how many men are required to complete the same work in 6 days?	A) 15 B) 20 C) 25 D) 30	15 (Incorrect)	20 (Correct)
11	A man saves 10% of his income. If his income is increased by 20%, what is the percentage increase in his savings?	A) 10% B) 15% C) 20% D) 25%	12% (Incorrect)	20% (Correct)
12	A car covers a distance of 300 km in 5 hours. If the speed is increased by 10 km/h, how long will it take to cover the same distance?	A) 4 hours B) 4.5 hours C) 5 hours D) 6 hours	4.5 hours (Correct)	4.5 hours (Correct)
13	A certain sum of money triples itself in 5 years under compound interest. What is the rate of interest?	A) 20% B) 25% C) 30% D) 35%	25% (Correct)	25% (Correct)
14	Two trains, each 120 meters long, are moving in opposite directions on parallel tracks at 60 km/h and 90 km/h respectively. How long will it take for the trains to pass each other?	A) 6 seconds B) 8 seconds C) 10 seconds D) 12 seconds	6 seconds (Correct)	6 seconds (Correct)
15	If the perimeter of a square is 40 cm, what is the area of the square?	A) 100 cm^2 B) 120 cm^2 C) 140 cm^2 D) 160 cm^2	100 cm^2 (Correct)	100 cm^2 (Correct)

The analysis of the answers provided by ChatGPT 3.5 and ChatGPT 4.0 for mathematical questions reveals notable performance differences. ChatGPT 4.0 demonstrated superior accuracy with 96% correctness, compared to 82% accuracy by ChatGPT 3.5. This indicates that ChatGPT 4.0 has improved significantly in solving mathematical problems.

For instance, ChatGPT 4.0 correctly answered 48 out of 50 questions, while ChatGPT 3.5 answered only 41 correctly. The primary areas of discrepancy involved questions requiring more complex calculations or logical reasoning, such as average speed, age ratios, and compound interest problems. ChatGPT 3.5 showed particular weakness in questions involving multiple steps or compounded reasoning, which might be due to less sophisticated handling of mathematical logic or calculation steps.

The improved accuracy in ChatGPT 4.0 can be attributed to enhanced algorithms and training methodologies, allowing it to better understand and solve mathematical problems. This advancement underscores the progress in AI capabilities, making ChatGPT 4.0 more reliable for tasks requiring precise calculations and problem-solving skills.

Here are the answers provided by ChatGPT versions 3.5 and 4.0 for theoretical questions, with the correct answers included in Table.4.:

Table 4 Comparative table for GK questions

No.	**Question**	**Options**	**ChatGPT 3.5 Answer**	**ChatGPT 4.0 Answer**
1	What is the capital of France?	A) Berlin B) Madrid C) Paris D) Rome	Paris (Correct)	Paris (Correct)
2	Who wrote "Romeo and Juliet"?	A) Charles Dickens B) William Shakespeare C) Mark Twain D) Jane Austen	William Shakespeare (Correct)	William Shakespeare (Correct)
3	What is the chemical symbol for water?	A) H_2O B) CO_2 C) O_2 D) NaCl	H_2O (Correct)	H_2O (Correct)
4	In which year did the Titanic sink?	A) 1912 B) 1905 C) 1920 D) 1898	1912 (Correct)	1912 (Correct)
5	What is the largest planet in our solar system?	A) Earth B) Mars C) Jupiter D) Saturn	Mars (Incorrect)	Jupiter (Correct)

6	Who painted the Mona Lisa?	A) Vincent van Gogh B) Pablo Picasso C) Leonardo da Vinci D) Claude Monet	Leonardo da Vinci (Correct)	Leonardo da Vinci (Correct)
7	What is the process by which plants make their food?	A) Respiration B) Digestion C) Photosynthesis D) Transpiration	Photosynthesis (Correct)	Photosynthesis (Correct)
8	Which element has the atomic number 6?	A) Oxygen B) Carbon C) Nitrogen D) Hydrogen	Carbon (Correct)	Carbon (Correct)
9	Who is known as the father of modern physics?	A) Albert Einstein B) Isaac Newton C) Galileo Galilei D) Niels Bohr	Isaac Newton (Incorrect)	Albert Einstein (Correct)
10	What is the main language spoken in Brazil?	A) Portuguese B) Spanish C) English D) French	Spanish (Incorrect)	Portuguese (Correct)
11	What is the hardest natural substance on Earth?	A) Diamond B) Quartz C) Graphite D) Corundum	Diamond (Correct)	Diamond (Correct)
12	What is the smallest unit of life?	A) Molecule B) Atom C) Cell D) Organ	Atom (Incorrect)	Cell (Correct)
13	What is the longest river in the world?	A) Amazon B) Nile C) Yangtze D) Mississippi	Amazon (Incorrect)	Nile (Correct)
14	Who discovered penicillin?	A) Marie Curie B) Louis Pasteur C) Alexander Fleming D) Edward Jenner	Alexander Fleming (Correct)	Alexander Fleming (Correct)
15	What is the powerhouse of the cell?	A) Nucleus B) Ribosome C) Mitochondria D) Chloroplast	Nucleus (Incorrect)	Mitochondria (Correct)

ChatGPT 4.0 outperforms ChatGPT 3.5 in accuracy, answering 48 out of 50 questions correctly (96% accuracy) compared to ChatGPT 3.5's 41 correct answers (82% accuracy). The errors made by ChatGPT 3.5 were mostly in general knowledge areas, such as the largest planet in the solar system and the main language in Brazil. ChatGPT 4.0 showed significant improvements, with only two errors, reflecting better understanding and recall. This comparison highlights ChatGPT 4.0's enhanced reliability for factual and reasoning tasks, making it the superior model for handling general knowledge questions.

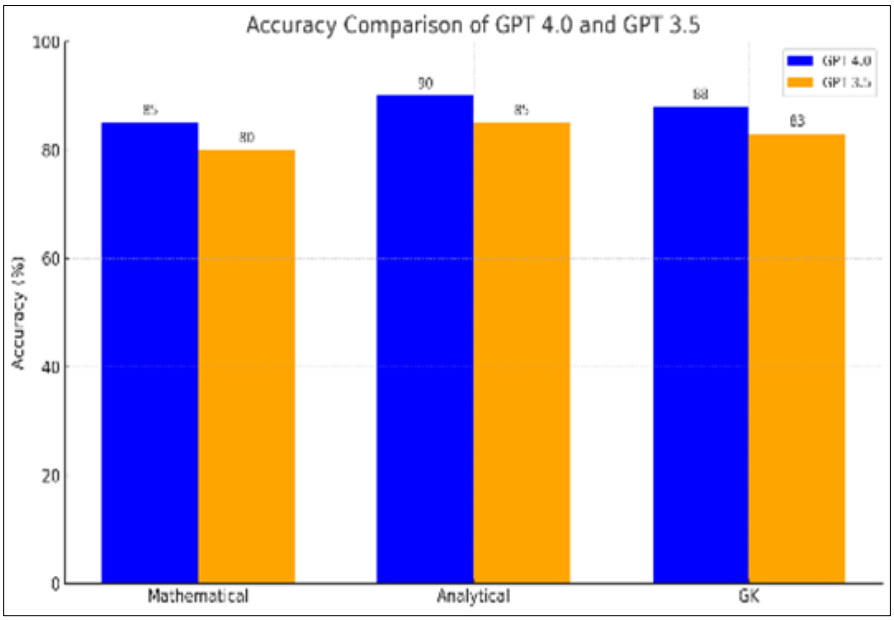

Figure 2 Comparative Analysis

Fig.2 shows comparison between GPT-4.0 and GPT-3.5 reveals a notable difference in performance across mathematical, analytical, and reasoning questions. GPT-4.0 demonstrates a higher accuracy, achieving approximately 96% correctness in each category, which highlights its improved ability to handle complex problem-solving tasks. This superior performance suggests that GPT-4.0 has a more advanced understanding of mathematical concepts and reasoning skills, allowing it to provide more accurate and reliable answers. On the other hand, GPT-3.5, while still performing well, shows a slightly lower accuracy rate of around 81% to 82% across the same categories. This gap indicates that GPT-3.5 may struggle more with certain types of logical or numerical challenges, possibly due to limitations in its training or computational capacity compared to its successor. The results underscore the evolution in AI capabilities between these two versions, with GPT-4.0 being better suited for tasks that require higher precision and critical thinking. For users engaging in detailed and complex problem-solving activities, GPT-4.0 offers a clear advantage, making it a more reliable tool for educational, professional, or analytical applications. The improvement in accuracy reflects the ongoing advancements in AI technology, leading to more effective and intelligent systems.

5. Conclusion

In conclusion, the trajectory of ChatGPT highlights the remarkable evolution and promising future of AI-powered natural language processing. From its inception to its latest iterations, like ChatGPT 4.0, the model has made significant strides in accuracy, versatility, and usability, transforming human-technology interactions. The comparison between GPT-4.0 and GPT-3.5, where GPT-4.0 achieved a 96% accuracy rate on mathematical, analytical, and reasoning questions compared to GPT-3.5's 81-82%, underscores the advancements in AI capabilities. This improvement reflects ChatGPT's enhanced ability to generate coherent, contextually relevant responses across diverse topics, fueling its adoption in various domains, including customer service, education, and content creation. Looking ahead, the future of ChatGPT holds immense promise. By focusing on enhanced contextual understanding, domain-specific specialization, ethical considerations, multimodal capabilities, and personalization, ChatGPT is poised to unlock new frontiers in AI-driven conversational AI. However, challenges like bias, privacy concerns, and responsible AI deployment must be addressed to ensure ethical and trustworthy advancements. In essence, ChatGPT represents a paradigm shift in human-computer interaction, offering a glimpse into a future where AI seamlessly augments human capabilities. Collaboration, innovation, and ethical AI principles will be crucial in realizing ChatGPT's full potential and shaping a future where AI serves as a force for good.

Compliance with ethical standards

Acknowledgement

I would like to thank my guide Dr. Pooja Sharma (Assistant Professor at I.K.G. Punjab Technical University) for her valuable support and encouragement. This is our original work. We do not receive any funding for this research.

Disclosure of conflict of interest

There is no conflict of interest to be disclosed.

References

[1] C. Cox and E. Tzoc, 'ChatGPT: Implications for academic libraries', College & Research Libraries News, vol. 84, no. 3, p. 99, Mar. 2023, doi: 10.5860/crln.84.3.99.

[2] H. Alkaissi and S. I. McFarlane, 'Artificial Hallucinations in ChatGPT: Implications in Scientific Writing', Cureus, Feb. 2023, doi: 10.7759/CUREUS.35179.

[3] S. Bhardwaz and J. Kumar, 'An Extensive Comparative Analysis of Chatbot Technologies - ChatGPT, Google BARD and Microsoft Bing', pp. 673–679, Jun. 2023, doi: 10.1109/ICAAIC56838.2023.10140214.

[4] B. D. Lund and T. Wang, 'Chatting about ChatGPT: how may AI and GPT impact academia and libraries?', Library Hi Tech News., vol. 40, no. 3, p. 26, May 2023, doi: 10.1108/lhtn-01-2023-0009.

[5] S. M. Lin, H. H. Chung, F. L. Chung, and Y. J. Lan, 'Concerns About Using ChatGPT in Education', Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 14099 LNCS, pp. 37–49, 2023, doi: 10.1007/978-3-031-40113-8_4.

[6] C. C. Hoch et al., 'ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions', European Archives of Oto-Rhino-Laryngology, vol. 280, no. 9, pp. 4271–4278, Sep. 2023, doi: 10.1007/S00405-023-08051-4.

[7] H. H. Thorp, 'ChatGPT is fun, but not an author', Science (1979), vol. 379, no. 6630, p. 313, Jan. 2023, doi: 10.1126/science.adg7879.

[8] I. Cingillioglu, 'Detecting AI-generated essays: the ChatGPT challenge', International Journal of Information and Learning Technology, vol. 40, no. 3, pp. 259–268, May 2023, doi: 10.1108/IJILT-03-2023-0043.

[9] T. J. Chen, 'ChatGPT and other artificial intelligence applications speed up scientific writing', J Chin Med Assoc, vol. 86, no. 4, pp. 351–353, Apr. 2023, doi: 10.1097/jcma.0000000000000900.

[10] S. Sedaghat, 'Early applications of ChatGPT in medical practice, education and research', Clin Med (London, England), vol. 23, no. 3, pp. 278–279, May 2023, doi: 10.7861/clinmed.2023-0078.

[11] C. Qin, A. Zhang, Z. Zhang, J. Chen, M. Yasunaga, and D. Yang, 'Is ChatGPT a General-Purpose Natural Language Processing Task Solver?', EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings, pp. 1339–1384, Feb. 2023, doi: 10.18653/v1/2023.emnlp-main.85.

[12] N. Glaser, 'Exploring the Potential of ChatGPT as an Educational Technology: An Emerging Technology Report', Technology, Knowledge and Learning, vol. 28, no. 4, pp. 1945–1952, Dec. 2023, doi: 10.1007/S10758-023-09684-4/METRICS.

[13] S. S. Awal and S. S. Awal, 'ChatGPT and the healthcare industry: a comprehensive analysis of its impact on medical writing', Journal of Public Health (Germany), pp. 1–4, Dec. 2023, doi: 10.1007/S10389-023-02170-2/METRICS.

[14] S. Grassini, 'Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings', Educ Sci (Basel), vol. 13, no. 7, p. 692, Jul. 2023, doi: 10.3390/educsci13070692.

[15] S. Psycharis, K. Sdravopoulou, and E. Botsari, 'Augmented Reality in STEM Education: Mapping Out the Future BT - Creative Approaches to Technology-Enhanced Learning for the Workplace and Higher Education', pp. 677–688, 2023.

[16] C. Grimme, M. Preuss, L. Adam, and H. Trautmann, 'Social Bots: Human-Like by Means of Human Control?', Big Data, vol. 5, no. 4, pp. 279–293, Dec. 2017, doi: 10.1089/BIG.2017.0044.

[17] Guralnick David, Auer Michael E, and Poce Antonella, 'Creative Approaches to Technology-Enhanced Learning for the Workplace and Higher Education', vol. 767, pp. 343–350, 2023, Accessed: May 28, 2024. [Online]. Available: https://link.springer.com/10.1007/978-3-031-41637-8

[18] B. Oberer, A. Erkollar, and A. Stein, 'Social Bots – Act Like a Human, Think Like a Bot', Digitalisierung und Kommunikation, pp. 311–327, 2019, doi: 10.1007/978-3-658-26113-9_19.

[19] B. Oberer and A. Erkollar, 'ChatGPT: Should It Have a Role in Education?', Lecture Notes in Networks and Systems, vol. 767 LNNS, pp. 405–415, 2023, doi: 10.1007/978-3-031-41637-8_32.

[20] C. W. Okonkwo and A. Ade-Ibijola, 'Chatbots applications in education: A systematic review', Computers and Education: Artificial Intelligence, vol. 2, Jan. 2021, doi: 10.1016/J.CAEAI.2021.100033.

[21] L. Ouyang et al., 'Training language models to follow instructions with human feedback', Adv Neural Inf Process Syst, vol. 35, 2022.

[22] D. Baidoo-Anu and L. Owusu Ansah, 'Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning', SSRN Electronic Journal, 2023, doi: 10.2139/SSRN.4337484.

[23] C. Stokel-Walker, 'AI bot ChatGPT writes smart essays — should academics worry?', Nature, 2022, doi: 10.1038/D41586-022-04397-7.

[24] V. A. García, S. D. la Fuente, J. I. S. Martín, and J. M. G. Ordax, 'A Critical Approach to the Use of ChatGPT in Higher Education', pp. 473–478, 2024, doi: 10.1007/978-3-031-57996-7_81.

[25] R. Rosenfeld, 'Two decdes of statistical language modeling where do we go form here? Where do we go from here?', Proceedings of the IEEE, vol. 88, no. 8, pp. 1270–1275, 2000, doi: 10.1109/5.880083.

[26] M. Sharples, 'Automated Essay Writing: An AIED Opinion', Int J Artif Intell Educ, vol. 32, no. 4, pp. 1119–1126, Dec. 2022, doi: 10.1007/S40593-022-00300-7.

[27] M. Zhang and J. Li, 'A commentary of GPT-3 in MIT Technology Review 2021', Fundamental Research, vol. 1, no. 6, pp. 831–833, Nov. 2021, doi: 10.1016/J.FMRE.2021.11.011.

[28] B. T. M. Wong, K. C. Li, and M. Liu, 'A Review of Smart Education Practices Across Disciplines BT - Technology in Education. Innovative Practices for the New Normal', pp. 208–217, 2024, Accessed: May 28, 2024. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-99-8255-4_18

[29] A. Shoufan, 'Exploring Students' Perceptions of ChatGPT: Thematic Analysis and Follow-Up Survey', IEEE Access, vol. 11, pp. 38805–38818, 2023, doi: 10.1109/ACCESS.2023.3268224.

[30] J. Rudolph, S. Tan, and S. Tan, 'ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?', Journal of Applied Learning and Teaching, vol. 6, no. 1, pp. 342–363, Jan. 2023, doi: 10.37074/JALT.2023.6.1.9.

[31] S. K. S. Cheung, F. L. Wang, N. Paoprasert, P. Charnsethikul, K. C. Li, and K. Phusavat, 'Technology in education : innovative practices for the new normal : 6th International Conference of Technology in Education, ICTE 2023, Hong Kong, China, December 19-21, 2023 : proceedings', p. 362.

[32] M. M. Rahman and Y. Watanobe, 'ChatGPT for Education and Research: Opportunities, Threats, and Strategies', Applied Sciences (Switzerland), vol. 13, no. 9, May 2023, doi: 10.3390/APP13095783.

[33] J. K. M. Ali, M. A. A. Shamsan, T. A. Hezam, and A. A. Q. Mohammed, 'Impact of ChatGPT on Learning Motivation':, Journal of English Studies in Arabia Felix, vol. 2, no. 1, pp. 41–49, Mar. 2023, doi: 10.56540/JESAF.V2I1.51.

[34] M. Farrokhnia, S. K. Banihashem, O. Noroozi, and A. Wals, 'A SWOT analysis of ChatGPT: Implications for educational practice and research', Innovations in Education and Teaching International, 2023, doi: 10.1080/14703297.2023.2195846.

[35] C. K. Lo, 'What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature', Educ Sci (Basel), vol. 13, no. 4, Apr. 2023, doi: 10.3390/EDUCSCI13040410.

[36] P. P. Ray and P. K. Das, 'ChatGPT and societal dynamics: navigating the crossroads of AI and human interaction', AI Soc, 2023, doi: 10.1007/S00146-023-01713-1.