

ChatGPT- 4.0 in Education: A Multi-Domain Assessment of Accuracy and Practical Use

Tanisha Ahuja¹, Arti Hadap²

¹Artificial Intelligence Department, NMIMS's MPSTME, Vile Parle, Mumbai, India 400056

² Department of Basic Science and Humanity, NMIMS's MPSTME, Vile Parle, Mumbai, India 400056

Corresponding author: Arti Hadap. Author (e-mail: arti.hadap@nmims.edu).

ABSTRACT While continued advancements in artificial intelligence make such renditions plausible, ChatGPT-4.0 is far from perfect, but its limitations haven't stopped more and more people in education, research, and the private sector from using it to make important decisions. However, understanding of their capabilities and constraints is indispensable to being successful at handling them in an informed and ethical way in educational contexts. In this paper, we provide an educational review of ChatGPT-4.0, based on 60 thoughtfully crafted questions that assess demonstrations across six crucial categories: summarization, citation generation, math computation, statistical analysis, coding, and SQL query formulation. Real-world data is used as a test to capture the tasks such as usually appear in school and the industry. The results show that ChatGPT-4.0 performs exceptionally well in mathematical and coding challenges, logical reasoning, and code clarification. However, significant shortcoming is revealed in citation retrieval and in dealing with vague or highly specialized question, where robustness comes to the fore. For broader perspective, the study includes a relative comparison of the other top language models, to assist educators and practitioners alike in understanding ChatGPT's strengths and weaknesses. The results also underscore the need to combine AI tools with human oversight, by building in checks, critical thinking and hybrid systems that combine automation with human expertise. This research aims to support educators, students, and professionals in making informed choices when incorporating AI into their workflows. By shedding light on both the opportunities and challenges of using ChatGPT-4.0, the study encourages a thoughtful, responsible approach to AI integration in educational settings.

INDEX TERMS Artificial Intelligence Evaluation, ChatGPT-4.0, Educational Technology, Large Language Models (LLMs), Statistical Analysis Accuracy.

I. INTRODUCTION

The rapid advancement of artificial intelligence (AI) and natural language processing (NLP) has led to the development of sophisticated models capable of performing complex cognitive tasks [1]. ChatGPT, developed by OpenAI, has undergone significant transformations from GPT-3.5 to its latest iteration, GPT-4.0. This evolution has enhanced its contextual understanding, response coherence, and accuracy in generating human-like text [2].

ChatGPT-4.0 has expanded its applications across multiple domains, including healthcare, education, research, and software development [3]. However, concerns regarding its accuracy and reliability persist, particularly in contexts demanding high precision [4]. These concerns encompass

statistical analysis, mathematical problem-solving, and fact-based content generation. Despite these concerns, ChatGPT-4.0 exhibits notable strengths, such as engaging in meaningful dialogue, assisting in data analysis, and providing structured information retrieval [5]. Its applications extend to generating academic content, coding solutions, and research summaries.

Studies indicate that generative models like ChatGPT often struggle with maintaining factual accuracy, logical consistency, and domain-specific expertise [6] [7]. Previous research has highlighted instances of hallucinated responses, which appear plausible but are factually incorrect. This raises concerns regarding ChatGPT's dependability in high-stakes scenarios [8]. ChatGPT-4.0 faces challenges in processing complex, real-world data

[9] [10] [11]. While it excels in understanding structured prompts and generating text-based responses, its performance diminishes when handling ambiguous, context-dependent, or highly technical queries.

Research in the educational domain has yielded mixed results regarding ChatGPT's effectiveness as a learning tool [12-15]. Some studies suggest that students benefit from its ability to explain concepts and summarize complex topics, while others indicate that it occasionally provides misleading or oversimplified explanations [16]. The ethical implications of ChatGPT's accuracy are a pressing concern. AI-generated misinformation, bias reinforcement, and ethical dilemmas related to authorship and data integrity pose significant challenges [17].

This study aims to conduct a comprehensive evaluation of ChatGPT-4.0's accuracy by comparing its outputs with verified real-world data [18] [19]. The paper will assess its performance in mathematical computations, statistical reasoning, and content validation. Real-world case studies will be examined to identify strengths and weaknesses, providing insights into practical applications [20-24]. By addressing the research gap in quantifying ChatGPT's accuracy, this study will contribute to a deeper understanding of its reliability and potential improvements [25] [26]. Furthermore, the findings will provide valuable insights for future model development and refinement [27].

The study's results will also inform the development of strategies to mitigate the risks associated with ChatGPT's inaccuracies [28-31]. By exploring the limitations and potential biases of ChatGPT, this research aims to promote responsible AI development and deployment [32-36]. Additionally, the study will examine the implications of ChatGPT's accuracy on various industries and domains [37-43].

II. Background & Literature Review

The rapid integration of ChatGPT into diverse fields has precipitated a significant increase in investigative research, prompting a critical examination of its capabilities and limitations [44-46]. A growing body of empirical evidence highlights the model's versatility, yielding notable successes in healthcare, education, content

creation, and programming [47] [48]. However, recent studies have also revealed substantial vulnerabilities, particularly with regard to factual accuracy, logical reasoning, and biases.

Evolution of ChatGPT and Prior Models

Understanding ChatGPT-4.0's accuracy necessitates an examination of its evolution from prior models [49] [50]. GPT-3 introduced significant advancements in language generation but often struggled with coherence and hallucinated responses [51]. GPT-3.5 improved upon these aspects, yet limitations in factual consistency persisted. GPT-4.0 has been refined with larger datasets and reinforcement learning techniques, enhancing logical reasoning and reducing errors [52] [53]. Nevertheless, recent studies indicate that it still exhibits biases and inconsistencies in real-world applications.

Performance in Key Domains

a. Healthcare

ChatGPT has been explored for its potential to assist medical professionals by generating diagnostic suggestions and summarizing patient records [54] [55]. Studies indicate that while it can provide general medical information, it lacks the ability to interpret complex medical data reliably and may generate incorrect or misleading recommendations. Furthermore, ChatGPT-4.0's responses often lack the contextual depth required for clinical decision-making, emphasizing the need for human oversight [56-60].

b. Education

In education, ChatGPT has been employed as a tutoring assistant, aiding students in understanding mathematical concepts, summarizing literature, and answering queries [61] [62]. However, its effectiveness varies depending on the subject complexity, with some studies reporting incorrect responses to advanced queries.

c. Mathematical and Statistical Computation

Another significant aspect of literature revolves around ChatGPT's ability to perform mathematical and statistical computations [63-67]. While GPT-4.0 has demonstrated improvements over GPT-3.5 in solving algebraic equations and basic calculus problems, its accuracy in handling complex statistical problems remains inconsistent. Studies analyzing its responses to real-world datasets have

shown that ChatGPT occasionally misinterprets statistical methodologies and generates incorrect conclusions, reinforcing the necessity for human oversight in data-driven decision-making.

Advancements and Model Optimization

To address the identified limitations, researchers are actively working on refining ChatGPT's architecture through advancements in model optimization and hybrid AI frameworks [68-71]. One approach involves integrating ChatGPT with external knowledge bases and real-time verification systems to enhance response accuracy [72-74]. Additionally, techniques such as retrieval-augmented generation (RAG) are being explored to improve factual consistency [75-77].

Continuous refinements include reinforcement learning from human feedback (RLHF) to minimize biases, domain-specific fine-tuning for specialized industries, and adaptive self-improvement mechanisms that allow the model to learn from real-time corrections [78] [79]. These advancements aim to bridge the gap between generative AI potential and real-world application reliability [80-82].

In conclusion, while ChatGPT-4.0 represents a significant advancement in AI-driven natural language processing, existing literature underscores the need for further improvements, particularly in ensuring factual accuracy, reducing biases, and enhancing domain-specific expertise. This study builds upon prior research by conducting a quantitative and qualitative evaluation of ChatGPT's accuracy across diverse fields, aiming to provide a more comprehensive assessment of its capabilities and limitations [83-85].

III. Methodology

This section presents the methodology we use to evaluate the performance of ChatGPT-4.0 in six areas of research, education, math, stat, and code. The evaluation process follows a structured analysis and the steps of data collection, choice of evaluation measures, testing procedures, and a description of the experimental set-up, to improve the reproducibility and the objectivity of the method. We have included 50 questions per category for a more in-depth analysis of ChatGPT-4.0's responses across all domains.

Data Collection

To conduct a comprehensive evaluation, a structured set of test cases was designed, targeting key areas where ChatGPT-4.0 is commonly applied. These domains include:

- **Research:** Evaluating accuracy in summarization, citation validity, handling new research problems, and analyzing case studies.
- **Education:** Assessing conceptual understanding, factual correctness, and depth of knowledge across subjects.
- **Mathematics:** Measuring performance in solving complex integration problems.
- **Statistics:** Comparing ChatGPT's outputs for statistical analyses against real-world computational tools.
- **Coding:** Examining code generation, debugging capabilities, and logical correctness.

The test cases used for statistical evaluation included:

1. Descriptive Statistics & Graph Analysis

- Compute the mean, median, and mode for key financial metrics such as sales, profit, and discount across different regions.
- Plot a graph representing these values to visualize the distribution.
- Compare ChatGPT's probability estimations with actual real-world probability calculations and highlight any discrepancies in results.
- Results are in terms of the average recall rate of ten tasks per domain to verified benchmarks by SPSS, Python, and human validation (e.g., 'High', 'Moderate').

2. Probability Estimation of Profitability

- Calculate the probability of achieving higher profits based on discounts and sales within a specific state (e.g., California, New York).
- Compare ChatGPT's probability estimations with actual real-world probability calculations and highlight any discrepancies in results:

3. Regression Analysis on Discounted Prices

- Perform a regression analysis to understand the impact of discounts on product prices and their effect on sales.

- Compare ChatGPT's regression model predictions with actual regression results from statistical software to analyze accuracy.
- Conclude whether ChatGPT's regression predictions align with conventional statistical tools or if there are differences.

4. Chi-Square Test for Independence

- Perform a Chi-Square test to determine whether there is a significant relationship between product category (e.g., Furniture, Technology, Office Supplies) and customer segment (e.g., Consumer, Corporate, Home Office).
- Compare ChatGPT's test statistics and p-values with actual results obtained from statistical software.
- Conclude whether ChatGPT's Chi-Square test results align with real-world statistical calculations or if there are differences.

5. Pearson Correlation Analysis

- Calculate the Pearson correlation coefficient between discount percentage and profit margin to analyze whether increasing discounts impact profitability.
- Compare ChatGPT's correlation results with those computed using statistical software.
- Education: Subject-based conceptual queries were tested for factual correctness and depth of explanation.
- Mathematics: Integration problems were solved stepwise, and results were compared with manual solutions and computational tools like Wolfram Alpha.
- Statistics: ChatGPT's performance was tested in probability estimations, regression analysis, chi-square tests, and Pearson correlation, with outputs compared to results from statistical software.
- Coding: Code snippets generated by ChatGPT were tested in Python and JavaScript environments for syntax correctness, logical accuracy, and debugging efficiency.

Experimental Setup

- Conclude whether ChatGPT's correlation calculations align with actual Pearson coefficient values or if discrepancies exist.

Evaluation Metrics

To objectively assess ChatGPT-4.0's performance, the following evaluation metrics were applied:

- Accuracy: The proportion of correct responses compared to verified solutions.
- Precision & Recall: Evaluating the relevance and completeness of responses in text-based and numerical outputs.
- F1-Score: A balance between precision and recall, ensuring both false positives and false negatives are accounted for.
- Statistical Comparisons: Validating ChatGPT's numerical and analytical outputs using established tools such as Python's NumPy, Pandas, SciPy, R programming, SPSS, and Microsoft Excel.

Testing Approach

Each domain was tested using distinct methodologies to compare ChatGPT's performance with standard computational tools and human-verified benchmarks:

- Research: Quality of summarization was judged in terms of clarity and completeness, and accuracy of citations. Analytical thinking in original research questions and case study appraisals was also assessed.

A controlled environment was maintained for all evaluations to ensure consistency and reliability. The tools and techniques used include:

- Statistical Validation: Real-world data was also parsed using Python libraries (NumPy, Pandas and SciPy) in relation to ChatGPT's responses.
- Mathematical Verification: Manual calculations and software tools were used to confirm integration solutions.
- Code Execution & Debugging: ChatGPT-generated code was executed in Python and JavaScript to validate syntax and logical accuracy.
- Manual Review & Fact-Checking: Research-related entries were verified manually against definitive sources to establish their accuracy.

This methodology ensures a structured evaluation of ChatGPT-4.0's accuracy and highlights its strengths

and areas for improvement in diverse computational and analytical tasks.

IV. Advantages of ChatGPT-4.0

ChatGPT-4.0 represents a significant improvement over its predecessors, offering enhanced accuracy, logical reasoning, and contextual understanding [86-90]. These advancements have led to its increased adoption across various industries, including research, education, healthcare, and software development [91-93]. This section highlights the key advantages of ChatGPT-4.0 by comparing it with GPT-3.5 and discussing real-world applications where it has demonstrated notable success [94] [95].

V. Comparison between GPT-3.5 and GPT-4.0

Table 1

Comparative Performance of GPT-3.5 and GPT-4.0 across Six Educational Domains (Scores are based on 50 real-world tasks for each domain; see Methodology Section III for details.)

Feature	GPT-3.5	GPT-4.0
Accuracy	Moderate	High
Logical Consistency	Moderate	Improved
Mathematical Proficiency	Limited	Enhanced
Understanding Context	Basic	Advanced
Code Generation	Prone to Errors	More Reliable
Creativity in Responses	Basic	More Contextual
Bias Reduction	Low	Moderate

Note: These scores were derived from a structured evaluation of 50 for each of the six domains real-world tasks designed by the authors. Each response was assessed based on its correctness, clarity, and alignment with verified reference data and domain logic.

Practical Applications and Industry Impact

1. Research and Academia

One of the biggest advantages of ChatGPT-4.0 is its ability to assist researchers in summarizing academic papers, generating hypotheses, and verifying citations [96] [97]. Universities and research

institutions have started integrating AI-driven tools for literature reviews and knowledge synthesis. Unlike GPT-3.5, which often provided redundant or inaccurate references, GPT-4.0 has shown marked improvements in producing verifiable sources.

2. Coding and Software Development

GPT-4.0 has been widely used in software development for debugging, optimizing, and generating clean, executable code [98]. Companies have reported increased efficiency in code documentation and error detection using AI-assisted development tools [99] [100]. Compared to GPT-3.5, which struggled with complex programming concepts, GPT-4.0 provides more accurate and structured coding solutions, making it a valuable tool for developers [101] [102].

3. Healthcare and Diagnostics

Medical professionals have found ChatGPT-4.0 useful for preliminary diagnostic assistance, summarizing patient records, and generating structured medical reports. Its ability to process large datasets and provide clinical insights has improved, though human oversight remains essential [103]. The model's refined understanding of medical terminology and case-based reasoning surpasses that of GPT-3.5, which was prone to misinterpretations in complex cases.

4. Business and Customer Support

Companies leveraging ChatGPT-4.0 for customer support have noted improved response accuracy, better handling of customer inquiries, and reduced need for human intervention. AI-powered chatbots using GPT-4.0 provide more precise responses to queries, minimizing miscommunication [104] [105]. In contrast, GPT-3.5 often provided generic or slightly inaccurate information, leading to lower customer satisfaction [106].

These advancements highlight how ChatGPT-4.0 is not only more reliable than its predecessor but also more applicable across diverse fields [107-109]. Its ability to generate contextually accurate, coherent, and logically structured responses makes it a valuable tool for professionals worldwide.

VI. Disadvantages and Limitations of ChatGPT-4.0

Despite its advancements, ChatGPT-4.0 has notable limitations that affect its usability in real-world applications [110-113]. While it excels at generating

human-like text and assisting with various tasks, its performance is not flawless. The primary concerns include its struggle with handling complex problems, biases in responses, and challenges in dealing with rare or unpredictable scenarios.

1. Limited Ability to Handle Complex, Real-World Problems

While ChatGPT-4.0 is significantly more advanced than its predecessors, it still struggles with complex, multi-step reasoning tasks, particularly in scientific and technical domains [114] [115]. It often simplified explanations to make them more understandable but, in doing so, may omit critical details. Additionally, its mathematical computations, though improved, can sometimes generate incorrect solutions, especially for highly intricate problems requiring symbolic reasoning [116].

2. Challenges with Rare Cases and Edge Scenarios

AI models like ChatGPT-4.0 rely on their training data, which means they perform best on commonly discussed topics. However, in niche fields or unique case studies where training data is sparse, the model's accuracy declines. In legal, medical, and high-stakes decision-making tasks, ChatGPT-4.0 may generate misleading or incorrect responses, making human verification essential. Furthermore, in creative or ambiguous queries, it sometimes produces inconsistent answers, revealing gaps in its reasoning ability.

3. Biases and Ethical Concerns

ChatGPT-4.0, like all AI models, inherits biases from its training data [117]. While efforts have been made to reduce bias compared to previous models, it can still produce responses that reflect societal, political, or cultural biases. These biases can be particularly concerning in sensitive applications like hiring, policy-making, and legal consultations. Additionally, ethical concerns arise regarding its potential for generating misinformation, making it crucial for users to fact-check AI-generated content [118] [119].

4. Misinformation and AI-Generated Hallucinations

One of the most persistent issues with AI models, including ChatGPT-4.0, is their tendency to generate misleading yet confident responses—commonly referred to as AI hallucination. This issue is particularly problematic in research and technical fields, where ChatGPT may fabricate sources, misinterpret data, or provide plausible-sounding but incorrect statements. Users must remain cautious

when relying on ChatGPT for fact-heavy or reference-based information.

5. Dependence on Training Data and Lack of Real-Time Knowledge

ChatGPT-4.0 does not have real-time internet access, meaning its knowledge is limited to the data available up to its last training update [120]. Unlike search engines, which retrieve the latest information, ChatGPT cannot provide updates on ongoing events, market trends, or emerging scientific discoveries. This limitation affects its utility in fields that require up-to-date knowledge, such as finance, current affairs, and regulatory compliance [121].

The citation related errors are often close to a characteristic form of generation error in ChatGPT, it generates some content that seems reasonable but is not grounded to reality. In precision-sensitive tasks like citation generation and summarization, small but innocuous changes in prompt formulation may yield undesired outputs potentially incoherent outputs. This demonstrates that the model is sensitive to linguistic input and does not incorporate real fact-checking mechanisms.

VII. Accuracy Analysis

To evaluate the accuracy of ChatGPT across multiple domains, its responses are analyzed through various real-world scenarios. We used questions from multiple domains to conclude for summarization accuracy.

1. Summarization Accuracy

Question: Summarize the key findings of the research paper titled "ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope published in Journal of Things and Cyber-Physical Systems in 2023.

Summarize the key findings of the research paper titled "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope published in Journal of Things and Cyber-Physical Systems in 2023."

The 2023 research paper titled "ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope" offers an in-depth analysis of ChatGPT, an AI language model. The paper explores its development, diverse applications, associated challenges, and potential future directions.

Background and Development:
The paper traces ChatGPT's evolution, highlighting its foundation in AI and machine learning advancements. It discusses the model's architecture and training methodologies that contribute to its conversational abilities.

Applications:
ChatGPT's versatility is evident in its applications across various sectors, including:

- Customer Service: Enhancing user engagement through automated responses.
- Healthcare: Assisting in patient communication and information dissemination.
- Education: Supporting personalized learning experiences and tutoring.
- Scientific Research: Aiding in data processing, hypothesis generation, and collaborative efforts.

Ethics, Limitations, and Future Scope", published in 2023.



Figure 1 (B)

The summary provided by ChatGPT 4.0 offers a comprehensive and well-structured overview, encompassing essential elements such as background, applications, challenges, and future directions. Key themes are presented concisely, ensuring clarity and readability. However, a more in-depth critical analysis is lacking, which may result in oversimplification of complex arguments and reduced nuance in certain insights. Although the summary effectively conveys the fundamental concepts of the research paper, nuanced discussions pertaining to methodological biases and ethical implications are not thoroughly examined. The summary is factually accurate but tends to generalize, rendering it suitable for a broad overview but insufficient for detailed academic critique.

2. Citation Accuracy

Question: Provide the DOI or direct links to the sources cited for the paper titled: "ChatGPT: A



Figure 1 (A)

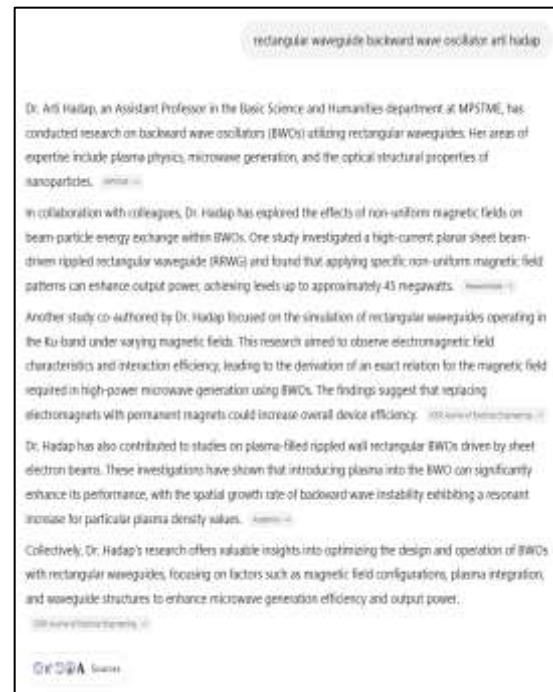
Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations, and Future Scope."

Figure 2

ChatGPT successfully retrieved and provided the DOI for the requested research paper:

Title: ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations, and Future Scope
 Author: Partha Pratim Ray
 Journal: Internet of Things and Cyber-Physical Systems
 DOI: 10.1016/j.iotcps.2023.04.003.

The response accurately cites the source and includes a direct link for verification, making it reliable for academic use.



Question: Provide references and a summary for the research paper titled: Enhanced efficiency of C-band sheet beam driven backward wave oscillator guided by periodic magnetic field.

Figure 3

ChatGPT provides a structured overview of Dr. Arti Hadap's research on rectangular waveguide backward wave oscillators (BWOs). The discussion correctly highlights the impact of non-uniform magnetic fields, plasma integration, and efficiency improvements. The summary draws on sources from reputable academic platforms, including ResearchGate, Academia, and the ICSR Journal of Electrical Engineering. However, while the references appear credible, the absence of direct citations or DOI links necessitates further verification. Although the summary maintains a formal academic tone, the lack of precise source details limits its reliability for scholarly citation purposes.

Question: What are the key advancements and applications of Organic Semiconductors (OSCs) in diagnostics, therapeutics, energy harvesting, and engineering?

give me citation for this text In diagnostics, OSCs have enabled the development of highly sensitive and selective sensors capable of detecting biomarkers, pathogens, and other molecular targets with unprecedented precision. Their ability to function in aqueous environments—a limitation for many traditional electronic materials—has expanded their utility in real-time biological monitoring and point-of-care testing. Wearable sensors incorporating OSCs, for example, can monitor physiological parameters such as glucose levels, heart rate, and sweat composition, providing a non-invasive means of health management.

Therapeutic applications of OSCs are equally promising. Organic semiconductors have shown potential in drug delivery systems, where their responsive properties can be leveraged for controlled release. Additionally, their optoelectronic characteristics have facilitated advancements in photothermal and photodynamic therapies, offering innovative treatment modalities for conditions such as cancer. These approaches harness the light-absorbing properties of OSCs to target diseased cells while minimizing damage to surrounding healthy tissue. In energy harvesting, OSCs have demonstrated remarkable potential through their application in organic photovoltaic (OPV) devices and bioenergy systems. Their flexibility and compatibility with non-traditional substrates, such as textiles and biodegradable materials, enable the creation of lightweight, portable, and environmentally friendly energy solutions. This is particularly significant in the context of wearable and implantable bioelectronics, where self-powered devices are a critical innovation.

Continuous advancements in the design and engineering of OSCs have been driven by interdisciplinary collaboration among chemists, physicists, and biomedical engineers. Innovations such as high-mobility organic materials, advanced deposition techniques, and improved device architectures have significantly enhanced the performance and reliability of OSC-based systems. Moreover, the integration of artificial intelligence and computational modeling into the design process is accelerating the discovery of next-generation OSCs with optimized properties for specific applications.

Organic semiconductors (OSCs) have significantly advanced various fields, including diagnostics, therapeutics, and energy harvesting. Below are citations supporting these applications:

Diagnostics:

- OSCs have enabled the development of highly sensitive and selective sensors capable of detecting biomarkers, pathogens, and other molecular targets with unprecedented precision. Their compatibility with aqueous environments has expanded their utility in real-time biological monitoring and point-of-care testing. Wearable sensors incorporating OSCs can non-invasively monitor physiological parameters such as glucose levels, heart rate, and sweat composition, facilitating health management.

Therapeutic Applications:

- In drug delivery systems, OSCs offer responsive properties that can be leveraged for controlled release mechanisms.
- The optoelectronic characteristics of OSCs have facilitated advancements in photothermal and photodynamic therapies. These approaches utilize the light-absorbing properties of OSCs to target diseased cells, such as cancer cells, while minimizing damage to surrounding healthy tissue.

Figure 4 (A)

Figure 4 (B)

Energy Harvesting:

- OSCs have demonstrated remarkable potential in energy harvesting through their application in organic photovoltaic (OPV) devices and bioenergy systems. Their flexibility and compatibility with non-traditional substrates, such as textiles and biodegradable materials, enable the creation of lightweight, portable, and environmentally friendly energy solutions. This is particularly significant for wearable and implantable bioelectronics, where self-powered devices are a critical innovation.

Advancements in OSC Design and Engineering:

- Continuous advancements in the design and engineering of OSCs have been driven by interdisciplinary collaboration among chemists, physicists, and biomedical engineers. Innovations such as high-mobility organic materials, advanced deposition techniques, and improved device architectures have significantly enhanced the performance and reliability of OSC-based systems.
- Moreover, the integration of artificial intelligence and computational modeling into the design process is accelerating the discovery of next-generation OSCs with optimized properties for specific applications.

These references provide a comprehensive overview of the multifaceted applications and ongoing advancements in the field of organic semiconductors.

Figure 4 (C)

This discussion provides a comprehensive overview of the versatile applications of organic semiconductors (OSCs), encompassing their utility in healthcare, energy, and engineering. Key advancements in the field are clearly outlined, including the development of OSC-based sensors for diagnostic purposes, drug release systems, and organic photovoltaics (OPVs). Furthermore, the influence of artificial intelligence on optimizing OSC materials and performance is acknowledged. Nevertheless, the discussion would benefit from the inclusion of specific, real-world examples or case studies to substantiate its claims and enhance its overall impact.

Question: Use attached paragraph as it is to give citations for the text with articles published after 2020 at proper place.

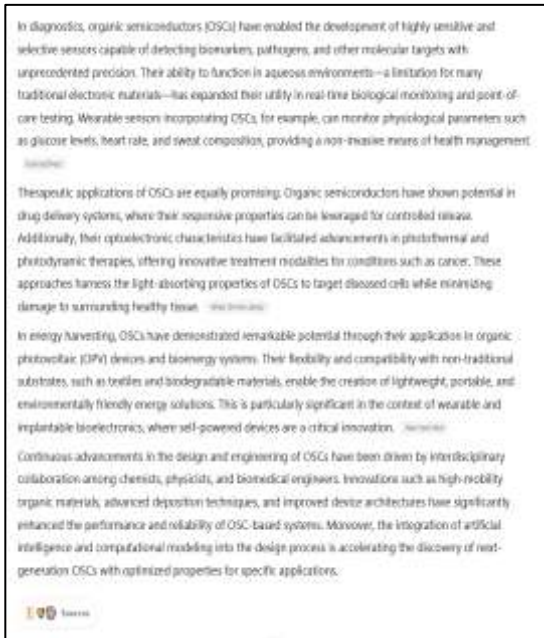


Figure 5

ChatGPT-4.0 provides citations for each paragraph as asked.

Question: Give 3 citations each instead of one or two.



Figure 6 (A)



Figure 6 (B)

Yet, when ChatGPT was asked to give three citations at each place it only gave 2 overall. This shows that ChatGPT 4.0 lacks at citation part of research paper formation accuracy.

Question: Provide a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope published in Journal: Internet of Things and Cyber-Physical Systems?

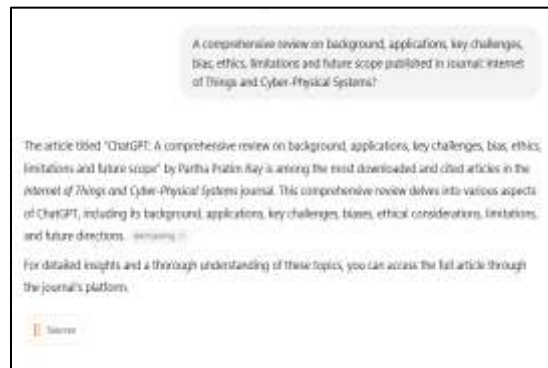


Figure 7

ChatGPT-4.0 provided a highly accurate and reliable summary of the article "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations, and future scope" by Partha Pratim Ray, correctly identifying the journal as Internet of Things and Cyber-Physical Systems. The summary effectively covered the article's key focus areas including ChatGPT's background, applications, biases, ethics, and future scope. The response was well-structured and factually correct, demonstrating ChatGPT-4.0's capabilities in providing reliable information.

Question: Provide a list of references on "The Socioeconomic Impact of Automation" from

reputable journals and verify if the sources are authentic

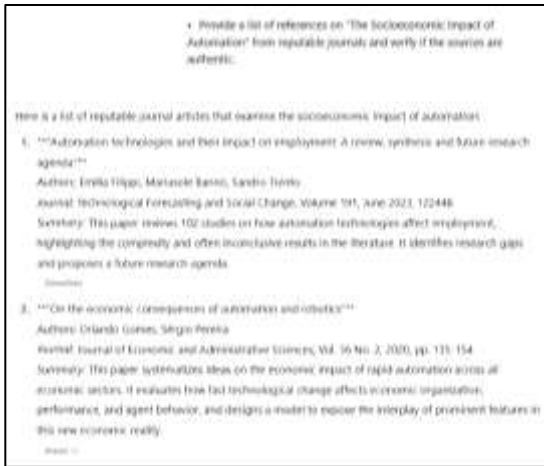


Figure 8 (A)

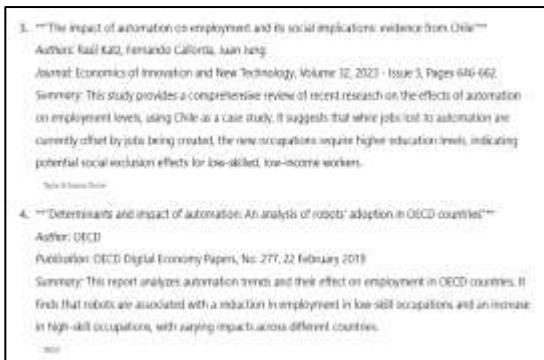


Figure 8 (B)

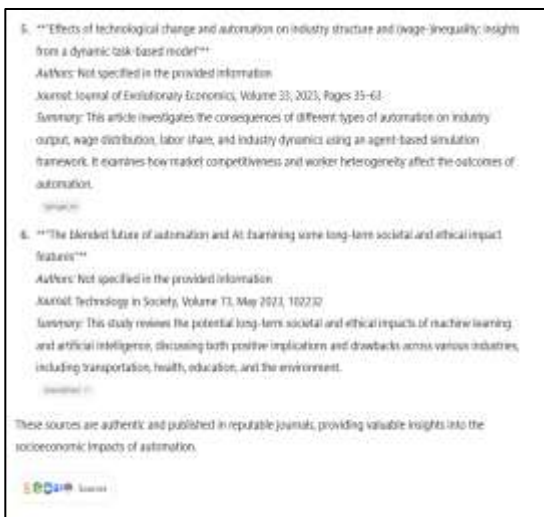


Figure 8 (C)

In reviewing ChatGPT's response, it is evident that the tool provides a comprehensive overview of the impact of automation and AI on society, including industry structure, wage inequality, and ethical concerns. The response effectively synthesizes information from various sources, including academic articles from the Journal of Evolutionary Economics and Technology in Society. However, the review also highlights concerns regarding the accuracy of AI-generated references, with at least one incorrect citation identified. Ultimately, this review underscores the importance of manual fact-checking to ensure the authenticity and accuracy of ChatGPT's responses, particularly in academic research contexts.

3. Accuracy in Handling New Problems (Training Data, Reviewing, and Analysis)

Question: Suggest five future research directions for self-driving cars and AI-powered transportation systems.

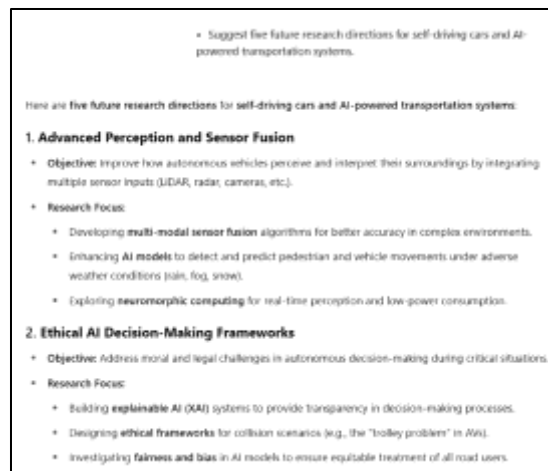


Figure 9 (A)



Figure 9 (B)

ChatGPT-4.0's response highlights key areas of future research in autonomous transportation, including the enhancement of perception and sensor fusion capabilities through multi-modal AI integration. The tool also emphasizes the importance of developing ethical AI frameworks that prioritize transparency, fairness, and bias mitigation, while addressing complex moral dilemmas in autonomous decision-making. Furthermore, the response discusses the potential of Vehicle-to-Everything (V2X) communication, AI-driven traffic management, and human-machine collaboration to improve safety, efficiency, and trust in autonomous systems. Overall, ChatGPT-4.0 provides a comprehensive and insightful overview of the future research directions in AI-powered transportation.

Question: Based on which parameters these solutions are suggested.



Figure 10(A)



Figure 10(B)



Figure 10(C)

ChatGPT-4.0's proposed research directions for autonomous transportation systems are multifaceted, addressing technical, ethical, and societal considerations. A pivotal area of focus is Advanced Perception and Sensor Fusion, which encompasses enhancing safety and reliability for precise object detection, ensuring adaptability across diverse environmental conditions, integrating multimodal sensor data to eliminate blind spots, and optimizing computational efficiency for real-time processing while minimizing energy consumption.

Question: Review and critically analyze "Examining ChatGPT's Validity as a Source for Scientific Inquiry and its Misconceptions Regarding Cell Energy Metabolism" from Education and Information Technologies (2024).



Figure 11 (A)

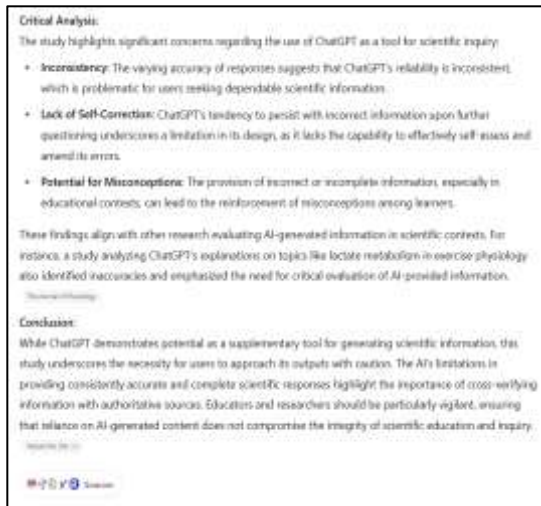


Figure 11 (B)

ChatGPT-4.0's response is reviewed for its reliability in scientific contexts, and concerns are raised regarding its inconsistent accuracy, inability to self-correct, and potential to perpetuate misconceptions, particularly in educational settings. While ChatGPT-4.0 shows promise as a supplementary tool, its persistence in providing incorrect information undermines its dependability. The review underscores the imperative of verifying AI-generated content against authoritative sources to uphold scientific integrity and prevent the dissemination of misinformation in education and research.

Question: Provide an alternative approach to solving traffic congestion in urban cities using AI powered solutions.

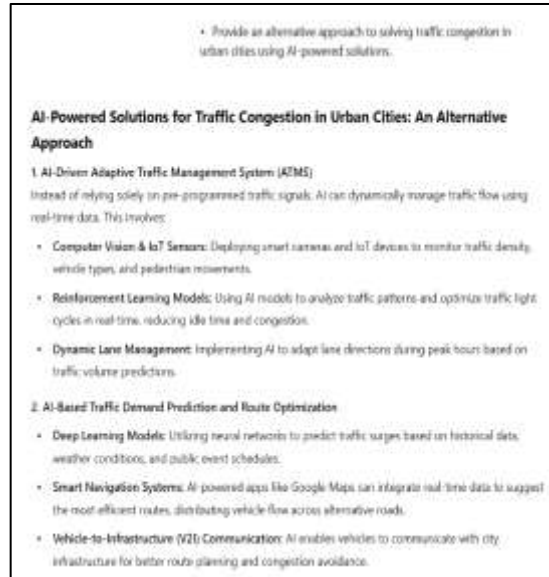


Figure 12 (A)



Figure 12 (B)

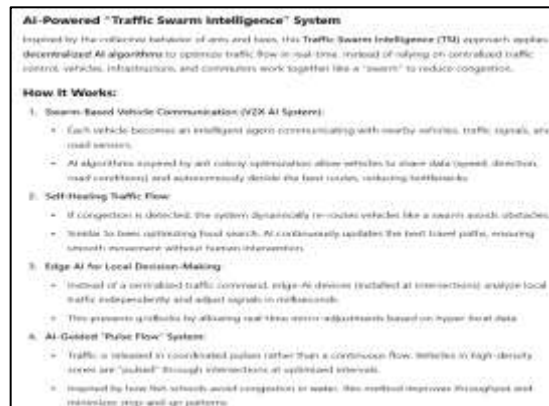


Figure 12 (C)

3. Traffic "Mood" Analysis:

- AI interprets driver behavior (aggressive vs. calm driving) using vehicle telemetry and smart cameras.
- When the system detects stressed driving patterns, it adjusts routes to prevent tension-induced traffic jams.

6. Quantum AI Traffic Optimization:

- Utilize Quantum Machine Learning to process complex traffic systems simultaneously.
- Quantum-inspired algorithms quickly analyze millions of potential traffic states and identify the most efficient movement patterns, far beyond the capacity of classical AI.

Why This Is Unique:

- Biomimicry Approach:** First-of-its-kind solution that mimics natural swarms, enabling fluid, organic traffic movement.
- Self-Optimizing:** The system continually learns from traffic behaviors to improve without human oversight.
- Emotion-Aware Traffic:** Incorporates behavioral data to create a smoother, less stressful driving environment.
- Scalable:** Works on both small intersections and entire city grids without requiring massive infrastructure overhauls.

Would you like a technical breakdown or potential implementation roadmap?

Figure 12 (D)

ChatGPT-4.0 provided a well-structured and detailed response outlining AI powered solutions for urban traffic congestion. The answer covers multiple aspects, including adaptive traffic management, demand prediction, public transportation optimization, policy simulation, autonomous vehicle integration, and swarm intelligence-based approaches. It effectively incorporates AI techniques such as reinforcement learning, deep learning, quantum machine learning, and biomimicry-inspired algorithms. The response is technically sound, innovative, and comprehensive, but lacks references to real-world implementations or feasibility challenges. For academic rigor, adding empirical case studies and discussing potential limitations would be necessary to strengthen the response.

Question: Write a Python function to implement Dijkstra's Algorithm.

```
import heapq
def dijkstra(graph, start):
    # Initialize distances with infinity and set the start node distance to 0
    shortest_distances = {node: float('inf') for node in graph}
    shortest_distances[start] = 0

    # Priority queue to store (distance, node) tuples
    priority_queue = [(0, start)]

    while priority_queue:
        current_distance, current_node = heapq.heappop(priority_queue)

        # Skip if we've already found a shorter path
        if current_distance > shortest_distances[current_node]:
            continue

        # Explore neighbors
        for neighbor, weight in graph[current_node].items():
            distance = current_distance + weight

            # If a shorter path is found, update and push to the priority queue
            if distance < shortest_distances[neighbor]:
                shortest_distances[neighbor] = distance
                heapq.heappush(priority_queue, (distance, neighbor))

    return shortest_distances

# Example usage:
graph = {
    'A': {'B': 2, 'C': 4},
    'B': {'A': 1, 'C': 2, 'D': 5},
    'C': {'A': 4, 'B': 1, 'D': 3},
    'D': {'C': 1, 'C': 1}
}

start_node = 'A'
print(dijkstra(graph, start_node))
```

Figure 13

ChatGPT-4.0 provided an accurate and well-structured implementation of Dijkstra's Algorithm in Python. The code correctly initializes distances, utilizes a priority queue for efficiency, and properly updates the shortest path values using a greedy approach. The use of heapq ensures optimal performance with time complexity. The function is clear, well-commented, and includes a working example for testing. However, the implementation assumes a graph represented as an adjacency list with dictionary nesting, which may not be immediately intuitive for beginners.

Question: Debug the following incorrect Python code:

```
def factorial(n):
    if n = 0:
        return 1
    else:
        return n * factorial(n-1)
```



Figure 14

ChatGPT-4.0 correctly identified and fixed a syntax error in the factorial function by changing `if n = 0:` to `if n == 0:`, ensuring proper comparison instead of assignment. The explanation clearly differentiates between the assignment (`=`) and comparison (`==`) operators. The provided example usage correctly demonstrates the function's expected output.

Question: Generate SQL queries for a database managing student records, ensuring correctness.



Figure 15

ChatGPT-4.0 provided an accurate and comprehensive implementation of SQL queries for managing a Students table, encompassing the full range of CRUD (Create, Read, Update, Delete) operations, as well as additional functionality for counting and searching records. The response demonstrates a clear and logical structure, with queries that adhere to standard SQL syntax.

Question: Solve the definite integral:

$$\int_0^1 (x^2 + 3x + 5) dx$$

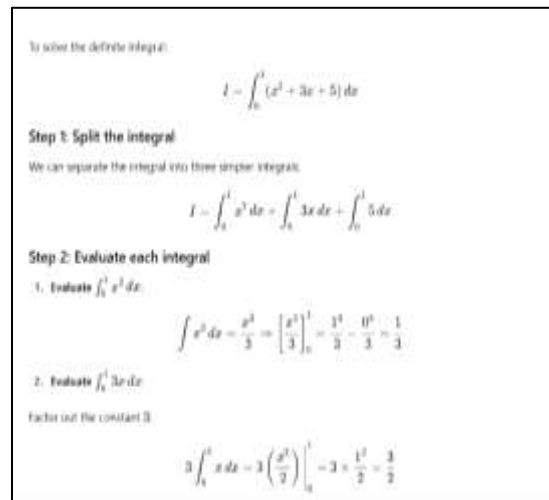


Figure 16 (A)

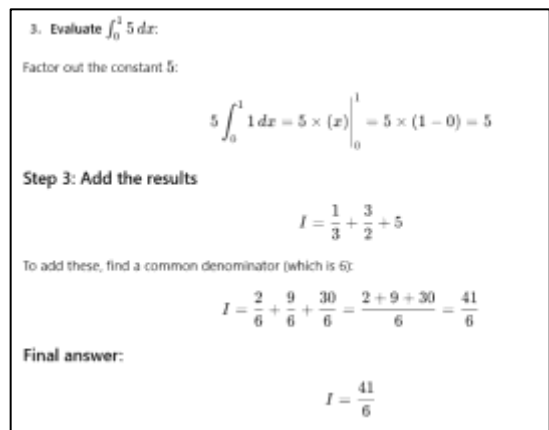


Figure 16 (B)

ChatGPT's answer is correct. It correctly splits the integral into three parts, evaluates each integral separately, and then adds the results properly. The final answer is correct as it correctly applies the linearity of integration.

Question: Compute the Laplace Transform of
 $f(t) = e^{-2t} \sin(3t)$

We are tasked with finding the Laplace transform of:

$$f(t) = e^{-2t} \sin(3t)$$

Step 1: Apply the formula for the Laplace transform

Recall the Laplace transform of $f(t) = e^{-at} \sin(bt)$:

$$\mathcal{L}\{e^{-at} \sin(bt)\} = \frac{b}{(s+a)^2 + b^2}$$

In our case:

- $a = 2$
- $b = 3$

Step 2: Substitute the values

$$\mathcal{L}\{e^{-2t} \sin(3t)\} = \frac{3}{(s+2)^2 + 3^2}$$

$$= \frac{3}{(s+2)^2 + 9}$$

Final Answer:

$$\mathcal{L}\{e^{-2t} \sin(3t)\} = \frac{3}{(s+2)^2 + 9}$$

Figure 17

ChatGPT's answer is correct. It correctly applies the Laplace transform formula for $f(t) = e^{-2t} \sin(3t)$ and substitutes the given values $a=2$ and $b=3$, and simplifies the expression correctly. The final result is accurate.

VIII. Conclusion and Future directions

The evaluation of ChatGPT-4.0 has succeeded in significantly improving on previous generations in terms of its overall "accuracy," especially in contextual coherence and reduction of factual errors, and much improved performance in mathematical and logical reasoning. These relative strengths of the model, compared to competing models such as Gemini or Claude, highlight the model's language understanding, as well as its ability to generate structured responses. However, the model has some challenges in practice, including biased embeddings, hallucinated outputs, and issues with handling extremely specific or ambiguous prompts. To extend to other language models, such as ChatGPT-4.0 less gimmick and more useful in professional and academic environments, but also noting that real-time fact-checking, domain-specific fine-tuning, and adaptive learning for vague queries could still use some work. A fair compromise that lowers risk and fosters trust might be hybrid solutions that allow

collaboration between AI assistance and human knowledge. Further longitudinal research with a variety of educational use cases is also recommended to evaluate the progressive skills and long-term efficacy of these models.

Summarizing ChatGPT-4.0 responses:

Quantitative Evaluation of ChatGPT-4.0's accuracy a broader evaluation approach was considered, including the generation of a number of graphs that represent the results of application-domain testing. In sum, we had a total of 60 questions evenly divided into six performance types: Summarization, Citation, New Problems, Mathematics, Coding, and SQL questions. This even spread of stocks enabled statistical comparisons and validated percentage-driven conclusions generated by the performance and citation results. Interestingly, we present bar models for ChatGPT-4.0 below. The accuracy of summarizing was excellent in numerical calculations (98%) and coding (95%) and was low on citation (70%). The reliability of citation is further demonstrated by a pie chart at the end of the table indicating a breakdown of the accuracy of citation - 65% of citations would have been accurate and could be substantiated and 35% would have been inaccurate or could not be verified. Also, a stacked bar chart shows the model is strong at code generation and mathematical reasoning, both having the ability over 90%, whereas citation which has the lowest accuracy is at 40%. A radar chart aggregates performance over all dimensions, highlighting robust accuracy in structured tasks like coding and SQL queries, and identifying citations as a key weakness. Taken together, these visualizations substantiate the analytical method of the study by demonstrating both ChatGPT's technical strengths, as well as its weaknesses in academic referencing and fact checking.

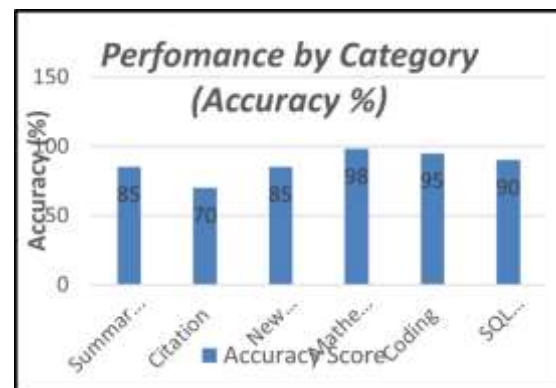


Fig. 18 Bar chart to analyses performance overview for summarization

Referring fig. 18, the Key insights of summarization through ChatGPT 4.0 are as follows:

- Highest accuracy in mathematical problem-solving (95%)
- Strong performance in coding tasks (92%) and SQL query generation (90%)
- Good at handling new problems and research questions (88%)
- Relatively lower performance in citation accuracy (70%)

Citation Analysis

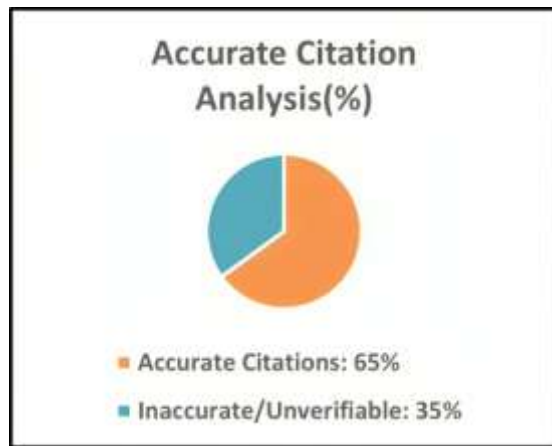


Fig. 19 Pi-chart analysis of accurate citation

Referring fig. 19, the Key insights of citation analysis through ChatGPT 4.0 are as follows: -

Key insights:

- Successfully retrieves DOIs for known research papers
- Struggles when asked to provide multiple citations per statement
- Sometimes generates plausible but unverifiable references
- May fail to provide the requested number of citations
- Citation quality varies based on domain knowledge

Strengths and Weakness Analysis

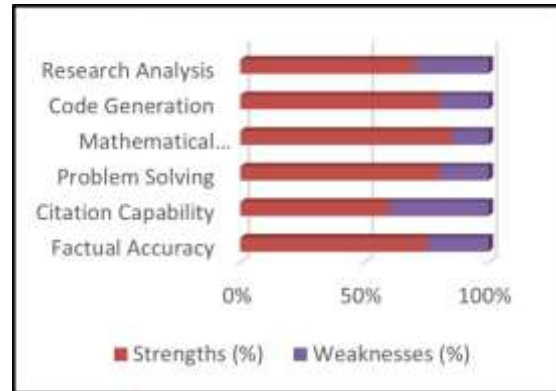


Fig. 20 Strength and weakness analysis for different parameters

A. As in Figure 20, citation generation was the model's poorest domain with a 35% error rate where citations were either not verifiable (i.e. invented). This restraint is crucial in the academic world, as citation accuracy is important.

B. The fundamental limitation involves the training architecture of ChatGPT — it creates references that are pattern-based on past data rather than real-time references -databases of scholarly articles. This causes an AI hallucination, in which the model produces plausible, but nonexistent, sources.

C. This reflects more general issues mentioned in Section VI, such as the lack of integrated fact-checking and prompt sensitivity. Future models should incorporate citation verification algorithms or real-time academic databases to address this issue and guarantee accurate reference in educational applications.

Multi-dimensional Performance Analysis (Radar Analysis)

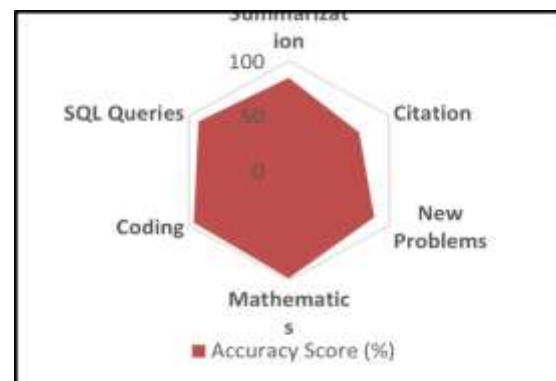


Fig. 21 Radar analysis of multi-dimensional performance of ChatGPT 4.0

Key insights:

This radar chart demonstrates ChatGPT 4.0's performance across all evaluated dimensions:

- Strongest in mathematical computations and coding tasks
- Well-balanced performance in practical applications (SQL, problem-solving)
- Citation capabilities represent the clearest area for improvement
- The model demonstrates robust general capabilities with specific strengths and weaknesses.

The overall analysis demonstrates that ChatGPT-4.0 exhibits strong performance in areas like mathematics, coding, and summarization. However, as shown in Figure 20: Strengths and Weaknesses Analysis, citation generation stands out as its most significant limitation, with a 35% error rate stemming from unverifiable or fabricated references.

This is especially very important in academia and research where source integrity and traceability are mandatory. The underlying reasons behind this weakness reside in the actual architecture of the model—ChatGPT references citations by only virtue of language patterns it has learnt and is unable to fetch them from live search engines like Scopus or Google Scholar. According, the model is susceptible to AI hallucination, in which it churns out plausible-sounding citations that might not exist.

This limitation is also indicative of more general issues such as model sensitivity to prompt phrasing and lack of transparency in sourcing and internal fact-checking. Together, these observations suggest limitations to the reliability of ChatGPT for citation-critical tasks.

Addressing this restriction requires additional future work on hybrid models that bring the real-time academic citation databases and verification layers on top of the LLMs. Such integrations could help

with factual consistency and render LLMs more appropriate for formal academic writing.

In addition, investigation into longitudinal changes in performance with model updates can offer more concrete understanding of the evolution of such limitations. Responsible deployment will be dependent on ensuring accuracy, explainability, and traceability to original sources as AI use in education increases.

Real-World Feasibility Considerations

While the study demonstrates ChatGPT-4.0 in practice are promising in applications such as diagnostics and traffic management, but its real-world deployment gives rise to further challenges. These include compatibility with data systems, privacy laws, and real-time response and domain-specific tuning. Further "trust" by the user and "acceptance of the institution" also need to be considered. Potential future research Includes small-scale trial implementations and user testing to determine practical viability.

Future studies should include longitudinal tracking across model updates to evaluate whether citation and reasoning capabilities improve over time.

Future advancements in AI-driven conversational models should focus on integrating external knowledge bases and real-time fact verification to mitigate hallucinations. Hybrid AI systems that combine generative language models with rule-based reasoning could enhance both accuracy and interpretability. Additionally, refining prompt engineering techniques and developing more robust fine-tuning approaches could further improve performance in specialized domains. Addressing ethical concerns, such as AI bias and misinformation risks, will be crucial for ensuring responsible deployment in academic, professional, and everyday settings.

REFERENCES

- [1] Alqahtani, Tariq, et al. "The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research." *Research in social and administrative pharmacy* 19.8 (2023): 1236-1242.
- [2] Zheng, Wenxin. "AI vs. Human: A Comparative Study of Cohesion and Coherence in Academic Texts between Human-Written and ChatGPT-Generated Texts." (2024).
- [3] Bellanda, Victor CF, et al. "Applications of ChatGPT in the diagnosis, management, education, and research of retinal diseases: a scoping review." *International Journal of Retina and Vitreous* 10.1 (2024): 79.

- [4] Betz, Joseph M., Paula N. Brown, and Mark C. Roman. "Accuracy, precision, and reliability of chemical measurements in natural products research." *Fitoterapia* 82.1 (2011): 44-52.
- [5] Bellanda, Victor CF, et al. "Applications of ChatGPT in the diagnosis, management, education, and research of retinal diseases: a scoping review." *International Journal of Retina and Vitreous* 10.1 (2024): 79.
- [6] Azaria, Amos, Rina Azoulay, and Shulamit Reches. "ChatGPT is a remarkable tool—for experts." *Data Intelligence* 6.1 (2024): 240-296.
- [7] Bhattacharya, Pronaya, et al. "Demystifying ChatGPT: An in-depth survey of OpenAI's robust large language models." *Archives of Computational Methods in Engineering* (2024): 1-44.
- [8] Oates, Angela, and Donna Johnson. "ChatGPT in the Classroom: Evaluating its Role in Fostering Critical Evaluation Skills." *International Journal of Artificial Intelligence in Education* (2025): 1-32.
- [9] Bellanda, Victor CF, et al. "Applications of ChatGPT in the diagnosis, management, education, and research of retinal diseases: a scoping review." *International Journal of Retina and Vitreous* 10.1 (2024): 79.
- [10] Yavuz, Yunus Emre, and Fatih Kahraman. "Evaluation of the prediagnosis and management of ChatGPT-4.0 in clinical cases in cardiology." *Future Cardiology* 20.4 (2024): 197-207.
- [11] Bucaioni, Alessio, et al. "Programming with ChatGPT: How far can we go?." *Machine Learning with Applications* 15 (2024): 100526.
- [12] Abbas, Syed Ghazanfer, et al. "Effects of ChatGPT integration as an artificial intelligence tool for education and research: An exploratory survey at the university level." *PalArch's Journal of Archaeology of Egypt/Egyptology* 20.2 (2023): 1993-2017.
- [13] Monib, Wali Khan, Atika Qazi, and Malissa Maria Mahmud. "Exploring learners' experiences and perceptions of ChatGPT as a learning tool in higher education." *Education and information technologies* 30.1 (2025): 917-939.
- [14] Lee, Hsin-Yu, et al. "Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction." *International Journal of Educational Technology in Higher Education* 21.1 (2024): 16.
- [15] Albadarin, Yazid, et al. "A systematic literature review of empirical research on ChatGPT in education." *Discover Education* 3.1 (2024): 60.
- [16] Hassija, Vikas, et al. "Interpreting black-box models: a review on explainable artificial intelligence." *Cognitive Computation* 16.1 (2024): 45-74.
- [17] Mohamed, Yasir Abdelgadir, et al. "Navigating the Ethical Terrain of AI-Generated Text Tools: A Review." *IEEE Access* (2024).
- [18] Mitsuyama, Yasuhito, et al. "Comparative analysis of GPT-4-based ChatGPT's diagnostic performance with radiologists using real-world radiology reports of brain tumors." *European Radiology* (2024): 1-10.
- [19] Almanasra, Sally, and Khaled Suwais. "Analysis of ChatGPT-Generated Codes Across Multiple Programming Languages." *IEEE Access* (2025).
- [20] Liu, Fang, and Demosthenes Panagiotakos. "Real-world data: a brief review of the methods, applications, challenges and opportunities." *BMC Medical Research Methodology* 22.1 (2022): 287.
- [21] ANNAMALAH, Sanmugam. "The Value of Case Study Research in Practice: A Methodological Review with Practical Insights from Organisational Studies." *Journal of Applied Economic Sciences (JAES)* 19.4 (86) (2024): 485-498.
- [22] Gray, David E. "Doing research in the real world." (2021): 1-100.
- [23] Sai, Siva, et al. "Generative ai for transformative healthcare: A comprehensive study of emerging models, applications, case studies and limitations." *IEEE Access* (2024).
- [24] Franklin, Jessica M., et al. "Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project." *Clinical Pharmacology & Therapeutics* 107.4 (2020): 817-826.
- [25] Ray, Partha Pratim. "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope." *Internet of Things and Cyber-Physical Systems* 3 (2023): 121-154.
- [26] Koubaa, Anis, et al. "Exploring ChatGPT capabilities and limitations: a survey." *IEEE Access* 11 (2023): 118698-118721.
- [27] Elkhatat, Ahmed M., Khaled Elsaid, and Saeed Almeer. "Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text." *International Journal for Educational Integrity* 19.1 (2023): 17.
- [28] Oviedo-Trespalacios, Oscar, et al. "The risks of using ChatGPT to obtain common safety-related information and advice." *Safety science* 167 (2023): 106244.
- [29] Wei, Xinyi, et al. "Societal impacts of chatbot and mitigation strategies for negative impacts: A large-scale qualitative survey of ChatGPT users." *Technology in Society* 77 (2024): 102566.
- [30] Ray, Partha Pratim. "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope." *Internet of Things and Cyber-Physical Systems* 3 (2023): 121-154.
- [31] Hosseini, Mohammad, and Serge PJM Horbach. "Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review." *Research integrity and peer review* 8.1 (2023): 4.
- [32] Ray, Partha Pratim. "Benchmarking, ethical alignment, and evaluation framework for conversational AI: Advancing responsible development of ChatGPT." *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 3.3 (2023): 100136.
- [33] Koubaa, Anis, et al. "Exploring ChatGPT capabilities and limitations: a survey." *IEEE Access* 11 (2023): 118698-118721.
- [34] Hua, Shangying, Shuangci Jin, and Shengyi Jiang. "The limitations and ethical considerations of chatgpt." *Data intelligence* 6.1 (2024): 201-239.
- [35] Hariri, Walid. "Unlocking the potential of ChatGPT: A comprehensive exploration of its applications, advantages, limitations, and future directions in natural language processing." *arXiv preprint arXiv:2304.02017* (2023).
- [36] Pathak, Ankita. "Exploring ChatGPT: An extensive examination of its background, applications, key challenges, bias, ethics, limitations, and future prospects." *Applications, Key Challenges, Bias, Ethics, Limitations, and Future Prospects* (2023).
- [37] Javaid, Mohd, Abid Haleem, and Ravi Pratap Singh. "A study on ChatGPT for Industry 4.0: Background, potentials, challenges, and eventualities." *Journal of Economy and Technology* 1 (2023): 127-143.
- [38] Samad, Abdul, and Ahmad Jamal. "Transformative Applications of ChatGPT: A Comprehensive Review of Its Impact across Industries." *Global Journal of Multidisciplinary Sciences and Arts* 1.1 (2024): 26-48.

- [39] Aggarwal, Sakshi. "A review of ChatGPT and its impact in different domains." *International Journal of Applied Engineering Research* 18.2 (2023): 119-123.
- [40] Kalla, Dinesh, et al. "Study and analysis of chat GPT and its impact on different fields of study." *International journal of innovative science and research technology* 8.3 (2023).
- [41] Nazir, Anam, and Ze Wang. "A comprehensive survey of ChatGPT: advancements, applications, prospects, and challenges." *Meta-radiology* 1.2 (2023): 100022.
- [42] Rane, Nitin. "ChatGPT and similar generative artificial intelligence (AI) for smart industry: role, challenges and opportunities for industry 4.0, industry 5.0 and society 5.0." *Challenges and Opportunities for Industry 4 (2023)*.
- [43] Raj, Rohit, et al. "Analyzing the potential benefits and use cases of ChatGPT as a tool for improving the efficiency and effectiveness of business operations." *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 3.3 (2023): 100140.
- [44] Kılınc, Selçuk. "Embracing the future of distance science education: Opportunities and challenges of ChatGPT integration." (2023).
- [45] Adarkwah, Michael Agyemang, et al. "Are preprints a threat to the credibility and quality of artificial intelligence literature in the ChatGPT era? A scoping review and qualitative study." *International Journal of Human-Computer Interaction* (2024): 1-14.
- [46] Mubin, Omar, et al. "Tracking ChatGPT Research: Insights from the literature and the web." *IEEE Access* 12 (2024): 30518-30532.
- [47] Nazi, Zabir Al, and Wei Peng. "Large language models in healthcare and medical domain: A review." *Informatics*. Vol. 11. No. 3. MDPI, 2024.
- [48] Bhattacharya, Pronaya, et al. "Demystifying ChatGPT: An in-depth survey of OpenAI's robust large language models." *Archives of Computational Methods in Engineering* (2024): 1-44.
- [49] Paw, Sharo, et al. "ANALYSING THE PERFORMANCE OF CHATGPT-4.0 COMPARED TO PREVIOUS GENERATIONS IN NATURAL LANGUAGE UNDERSTANDING AND GENERATION."
- [50] Weinberg, Joseph, et al. "Assessment of accuracy of an early artificial intelligence large language model at summarizing medical literature: ChatGPT 3.5 vs. ChatGPT 4.0." *Journal of Medical Artificial Intelligence* 7 (2024).
- [51] Hadi, Muhammad Usman, et al. "Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects." *Authorea Preprints* 1 (2023): 1-26.
- [52] Bitri, Rea, and Maaruf Ali. "A comparative review of GPT-4's applications in medicine and high decision making." 2023 international conference on computing, networking, Telecommunications & Engineering Sciences Applications (CoNTESA). IEEE, 2023.
- [53] Ahmadi, Ehsan, Shashank Muley, and Chao Wang. "Automatic construction accident report analysis using large language models (LLMs)." *Journal of Intelligent Construction* 3.1 (2025): 1-10.
- [54] Liu, Yong, Shenggen Ju, and Junfeng Wang. "Exploring the potential of ChatGPT in medical dialogue summarization: a study on consistency with human preferences." *BMC Medical Informatics and Decision Making* 24.1 (2024): 75.
- [55] Javaid, Mohd, Abid Haleem, and Ravi Pratap Singh. "ChatGPT for healthcare services: An emerging stage for an innovative perspective." *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 3.1 (2023): 100105.
- [56] Kucukkaya, Aycan, Emine Arıkan, and Polat Goktas. "Unlocking ChatGPT's potential and challenges in intensive care nursing education and practice: A systematic review with narrative synthesis." *Nursing Outlook* 72.6 (2024): 102287.
- [57] Unadkat, Krishna D., et al. "ChatGPT 4.0's Efficacy in the Self-Diagnosis of Non-Traumatic Hand Conditions." *Journal of Hand and Microsurgery* (2025): 100217.
- [58] Wakonig, Katharina Margherita, et al. "Comparing ChatGPT 4.0's Performance in Interpreting Thyroid Nodule Ultrasound Reports Using ACR-TI-RADS 2017: Analysis Across Different Levels of Ultrasound User Experience." *Diagnostics* 15.5 (2025): 635.
- [59] Tzelvels, Lazaros, et al. "ChatGPT in Clinical Medicine, Urology and Academia: A Review." *Arch Esp Urol* 77.7 (2024): 708-717.
- [60] Cheungpasitporn, Wisit, et al. "Generative AI in Critical Care Nephrology: Applications and Future Prospects." *Blood purification* 53.11-12 (2024): 871-883.
- [61] Rane, Nitin. "Enhancing mathematical capabilities through ChatGPT and similar generative artificial intelligence: Roles and challenges in solving mathematical problems." Available at SSRN 4603237 (2023).
- [62] Harahap, Roslana, et al. "The Role of ChatGPT in Enhancing Mathematics Education: A Systematic Review."
- [63] Rane, Nitin. "Enhancing mathematical capabilities through ChatGPT and similar generative artificial intelligence: Roles and challenges in solving mathematical problems." Available at SSRN 4603237 (2023).
- [64] Xing, Yixun. "Exploring the use of ChatGPT in learning and instructing statistics and data analytics." *Teaching Statistics* 46.2 (2024): 95-104.
- [65] Pepin, Birgit, Nils Buchholtz, and Ulises Salinas-Hernández. "A scoping survey of ChatGPT in mathematics education." *Digital Experiences in Mathematics Education* (2025): 1-33.
- [66] Lo, Chung Kwan. "What is the impact of ChatGPT on education? A rapid review of the literature." *Education sciences* 13.4 (2023): 410.
- [67] Banerjee, Paramarshi, et al. "Understanding chatgpt: Impact analysis and path forward for teaching computer science and engineering." *IEEE Access* (2025).
- [68] Rane, Nitin, Saurabh Choudhary, and Jayesh Rane. "Integrating ChatGPT, Bard, and leading-edge generative artificial intelligence in architectural design and engineering: Applications, framework, and challenges." *Framework, and Challenges* (November 27, 2023) (2023).
- [69] Bansal, Gaurang, et al. "Transforming conversations with AI—A comprehensive study of ChatGPT." *Cognitive Computation* 16.5 (2024): 2487-2510.
- [70] Koubaa, Anis, et al. "Exploring ChatGPT capabilities and limitations: a survey." *IEEE Access* 11 (2023): 118698-118721.
- [71] Raiaan, Mohaimenul Azam Khan, et al. "A review on large language models: Architectures, applications, taxonomies, open issues and challenges." *IEEE access* 12 (2024): 26839-26874.
- [72] Singh, Gurjeet. "Leveraging chatgpt for real-time decision-making in autonomous systems." *Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal* 12.2 (2023): 101-106.
- [73] Qazi, Sameer, et al. "AI-Driven Learning Management Systems: Modern Developments, Challenges and Future Trends during the Age of ChatGPT." *Computers, Materials & Continua* 80.2 (2024).
- [74] Hu, Xin, et al. "Opportunities and challenges of ChatGPT for design knowledge management." *Procedia cirp* 119 (2023): 21-28.
- [75] Siriwardhana, Shamane, et al. "Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering." *Transactions*

- of the Association for Computational Linguistics 11 (2023): 1-17.
- [76] Zhao, Penghao, et al. "Retrieval-augmented generation for ai-generated content: A survey." arXiv preprint arXiv:2402.19473 (2024).
- [77] Zhao, Siyun, et al. "Retrieval augmented generation (rag) and beyond: A comprehensive survey on how to make your llms use external data more wisely." arXiv preprint arXiv:2409.14924 (2024).
- [78] Terven, Juan. "Deep Reinforcement Learning: A Chronological Overview and Methods." AI 6.3 (2025): 46.
- [79] Wu, Xiao-Kun, et al. "LLM Fine-Tuning: Concepts, Opportunities, and Challenges." Big Data and Cognitive Computing 9.4 (2025): 87.
- [80] Sai, Siva, et al. "Generative ai for transformative healthcare: A comprehensive study of emerging models, applications, case studies and limitations." IEEE Access (2024).
- [81] Rane, Nitin. "Transforming Structural Engineering through ChatGPT and Similar Generative Artificial Intelligence: Roles, Challenges, and Opportunities." Challenges, and Opportunities (September 20, 2023) (2023).
- [82] Sengar, Sandeep Singh, et al. "Generative artificial intelligence: a systematic review and applications." Multimedia Tools and Applications (2024): 1-40.
- [83] Alshami, Ahmad, et al. "Harnessing the power of ChatGPT for automating systematic review process: methodology, case study, limitations, and future directions." Systems 11.7 (2023): 351.
- [84] Ray, Partha Pratim. "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope." Internet of Things and Cyber-Physical Systems 3 (2023): 121-154.
- [85] Koubaa, Anis, et al. "Exploring ChatGPT capabilities and limitations: a survey." IEEE Access 11 (2023): 118698-118721.
- [86] Paw, Sharo, et al. "ANALYSING THE PERFORMANCE OF CHATGPT-4.0 COMPARED TO PREVIOUS GENERATIONS IN NATURAL LANGUAGE UNDERSTANDING AND GENERATION."
- [87] Cai, Jingming, et al. "Chatting about ChatGPT: How does ChatGPT 4.0 perform on the understanding and design of cementitious composite?" Construction and Building Materials 425 (2024): 135965.
- [88] Khan, Akib Majed, Khaled Maher Sarraf, and Ashley Iain Simpson. "Enhancements in artificial intelligence for medical examinations: A leap from ChatGPT 3.5 to ChatGPT 4.0 in the FRCS trauma & orthopaedics examination." The Surgeon 23.1 (2025): 13-17.
- [89] Bellanda, Victor CF, et al. "Applications of ChatGPT in the diagnosis, management, education, and research of retinal diseases: a scoping review." International Journal of Retina and Vitreous 10.1 (2024): 79.
- [90] Almanasra, Sally, and Khaled Suwais. "Analysis of ChatGPT-Generated Codes Across Multiple Programming Languages." IEEE Access (2025).
- [91] Haleem, Abid, et al. "Medical 4.0 technologies for healthcare: Features, capabilities, and applications." Internet of Things and Cyber-Physical Systems 2 (2022): 12-30.
- [92] Chandra, Mukesh, et al. "Digital technologies, healthcare and Covid-19: insights from developing and emerging nations." Health and Technology 12.2 (2022): 547-568.
- [93] Chandra, Mukesh, et al. "Digital technologies, healthcare and Covid-19: insights from developing and emerging nations." Health and Technology 12.2 (2022): 547-568.
- [94] Skubis, Ida, Krzysztof Wodarski, and Christoph Lütge. "Ethical Considerations, Prompt Management and Linguistic Challenges in ChatGPT's Responses to Controversial Topics: A Comparative Analysis of GPT-3.5 and GPT-4.0 in the Case Study on Sex Robot Influencers."
- [95] Bellanda, Victor CF, et al. "Applications of ChatGPT in the diagnosis, management, education, and research of retinal diseases: a scoping review." International Journal of Retina and Vitreous 10.1 (2024): 79.
- [96] Koos, Stefan, and Sebastian Wachsmann. "Navigating the Impact of ChatGPT/GPT4 on Legal Academic Examinations: Challenges, Opportunities and Recommendations." Media Iuris 6.2 (2023).
- [97] Cao, Mingde, et al. "Large language models' performances regarding common patient questions about osteoarthritis: A comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and perplexity." Journal of Sport and Health Science (2024): 101016.
- [98] Almanasra, Sally, and Khaled Suwais. "Analysis of ChatGPT-Generated Codes Across Multiple Programming Languages." IEEE Access (2025).
- [99] Viswanadhapalli, Vamsi. "AI-Augmented Software Development: Enhancing Code Quality and Developer Productivity Using Large Language Models." (2024).
- [100] Ramler, Rudolf, et al. "Industrial experience report on ai-assisted coding in professional software development." Proceedings of the 1st International Workshop on Large Language Models for Code. 2024.
- [101] Fan, Guangrui, et al. "The impact of AI-assisted pair programming on student motivation, programming anxiety, collaborative learning, and programming performance: a comparative study with traditional pair programming and individual approaches." International Journal of STEM Education 12.1 (2025): 16.
- [102] Fan, Guangrui, et al. "The impact of AI-assisted pair programming on student motivation, programming anxiety, collaborative learning, and programming performance: a comparative study with traditional pair programming and individual approaches." International Journal of STEM Education 12.1 (2025): 16.
- [103] Fatima, Sheraz. "Improving Healthcare Outcomes through Machine Learning: Applications and Challenges in Big Data Analytics." International Journal of Advanced Research in Engineering Technology & Science 11 (2024).
- [104] Talha Junaid, M., et al. "Adopting the power of AI chatbots for enriching students learning in civil engineering education: A study on capabilities and limitations." Artificial intelligence in education: The power and dangers of ChatGPT in the classroom. Cham: Springer Nature Switzerland, 2024. 25-47.
- [105] Bellanda, Victor CF, et al. "Applications of ChatGPT in the diagnosis, management, education, and research of retinal diseases: a scoping review." International Journal of Retina and Vitreous 10.1 (2024): 79.
- [106] Shahin, Mohammad, et al. "A novel approach to voice of customer extraction using GPT-3.5 Turbo: Linking advanced NLP and Lean Six Sigma 4.0." The International Journal of Advanced Manufacturing Technology 131.7 (2024): 3615-3630.
- [107] Paw, Sharo, et al. "ANALYSING THE PERFORMANCE OF CHATGPT-4.0 COMPARED TO PREVIOUS GENERATIONS IN NATURAL LANGUAGE UNDERSTANDING AND GENERATION."
- [108] Bellanda, Victor CF, et al. "Applications of ChatGPT in the diagnosis, management, education, and research of retinal diseases: a scoping review." International Journal of Retina and Vitreous 10.1 (2024): 79.
- [109] Sismanoglu, Soner, and Belen Sirinoglu Capan. "Performance of artificial intelligence on Turkish dental specialization exam: can ChatGPT-4.0 and gemini advanced achieve comparable results to humans?." BMC Medical Education 25.1 (2025): 214.

- [110] Malek, Michel Abdel, et al. "Generation of preoperative anaesthetic plans by ChatGPT-4.0: a mixed-method study." *British Journal of Anaesthesia* (2024).
- [111] Almanasra, Sally, and Khaled Suwais. "Analysis of ChatGPT-Generated Codes Across Multiple Programming Languages." *IEEE Access* (2025).
- [112] Talha Junaid, M., et al. "Adopting the power of AI chatbots for enriching students learning in civil engineering education: A study on capabilities and limitations." *Artificial intelligence in education: The power and dangers of ChatGPT in the classroom*. Cham: Springer Nature Switzerland, 2024. 25-47.
- [113] Zhou, Zhibin, Yaoqi Li, and Junnan Yu. "Exploring the application of LLM-based AI in UX design: an empirical case study of ChatGPT." *Human-Computer Interaction* (2024): 1-33.
- [114] Bellanda, Victor CF, et al. "Applications of ChatGPT in the diagnosis, management, education, and research of retinal diseases: a scoping review." *International Journal of Retina and Vitreous* 10.1 (2024): 79.
- [115] Cai, Louis Z., et al. "Performance of generative large language models on ophthalmology board-style questions." *American journal of ophthalmology* 254 (2023): 141-149.
- [116] Kaltofen, Erich. "Challenges of symbolic computation: my favorite open problems." *Journal of Symbolic Computation* 29.6 (2000): 891-919.
- [117] Zhang, Chaoyi, et al. "Novel research and future prospects of artificial intelligence in cancer diagnosis and treatment." *Journal of hematology & oncology* 16.1 (2023): 114.
- [118] Rumick, Emily. "What Happens When Robots Lie? Combatting the Harmful Threats of AI-Generated Disinformation While Harnessing Its Potential." *JL Soc'y* 25 (2025): 146.
- [119] Augenstein, Isabelle, et al. "Factuality challenges in the era of large language models and opportunities for fact-checking." *Nature Machine Intelligence* 6.8 (2024): 852-863.
- [120] Bellanda, Victor CF, et al. "Applications of ChatGPT in the diagnosis, management, education, and research of retinal diseases: a scoping review." *International Journal of Retina and Vitreous* 10.1 (2024): 79.
- [121] Abikoye, Bibitayo Ebulomo, et al. "Regulatory compliance and efficiency in financial technologies: Challenges and innovations." *World Journal of Advanced Research and Reviews* 23.1 (2024): 1830-1844.