IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

USING DIFFERENT AI CHATBOTS TO FIND DIFFERENCES IN ANSWER AND THEIR ACCURACY LEVELS IN MULTIPLE SUBJECT GENRES.

[1]Sivun Panda, [2]Vipul Singh, [3]Tumul Nigam, [4]Aryan Kumar, [5]Pratik Dash

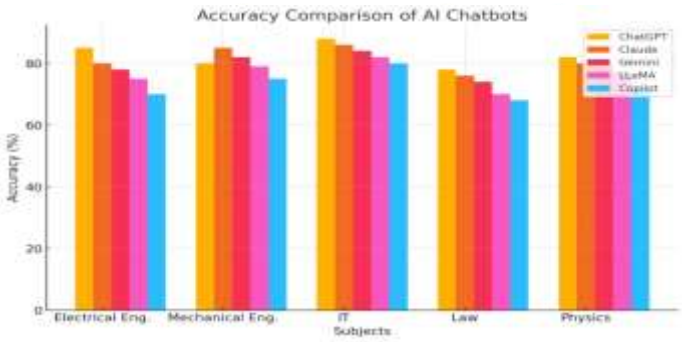
B.Tech Computer Science and Engineering, Minor in Economics and Finance.
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY, Bhubaneswar, Odisha, India

Abstract : **Artificial Intelligence (AI) chatbots are increasingly being used to provide responses across various academic and professional fields. This study evaluates the performance of five AI chatbots—Claude, Gemini, LLaMA, Copilot, and ChatGPT—by analysing the accuracy, numerical precision, and contextual understanding of their responses to a standardized questionnaire covering multiple disciplines. Through graphical analysis, including bar charts, heatmaps, and scatter plots, this research identifies the comparative strengths and weaknesses of these chatbots and their effectiveness in handling subject specific inquiries**

IndexTerms **– AI chatbot, Performance, analysis, etc.**

INTRODUCTION

AI chatbots serve as powerful tools for assisting in education and professional problem-solving, yet their performance varies significantly across different knowledge domains. The objective of this research is to systematically evaluate the responses of ChatGPT, Claude, Gemini, LLaMA, and Copilot, focusing on accuracy, numerical reasoning, and contextual depth. The study spans disciplines such as Electrical Engineering, Mechanical Engineering, Information Technology, Physics, and Law, identifying discrepancies and highlighting chatbot reliability in different fields.



LITERATURE REVIEW.

Existing research has analysed AI chatbot capabilities, demonstrating their strengths in theoretical explanations while also exposing their challenges in handling numerical and logical reasoning problems. AI-generated responses can vary due to differing training methodologies, data exposure, and model architectures. This study builds upon previous research by comparing multiple chatbots within a structured evaluation framework, assessing their accuracy, consistency, and subject-specific expertise.

METHODOLOGY.

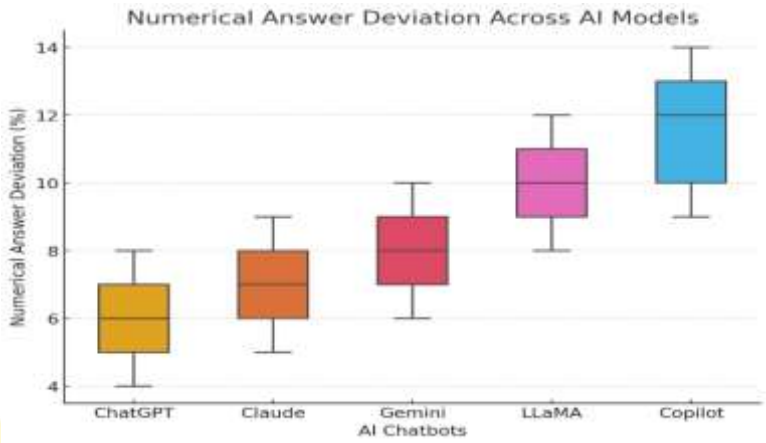
Acomprehensive questionnaire was designed to assess chatbot performance across multiple academic disciplines, including Electrical Engineering, Mechanical Engineering, IT, Physics, and Law. Each chatbot was tested with identical questions, and

responses were evaluated based on correctness, numerical accuracy, and contextual relevance. Data visualization techniques such as bar charts, heatmaps, scatter plots, and box plots were utilized to provide insights into performance variations.

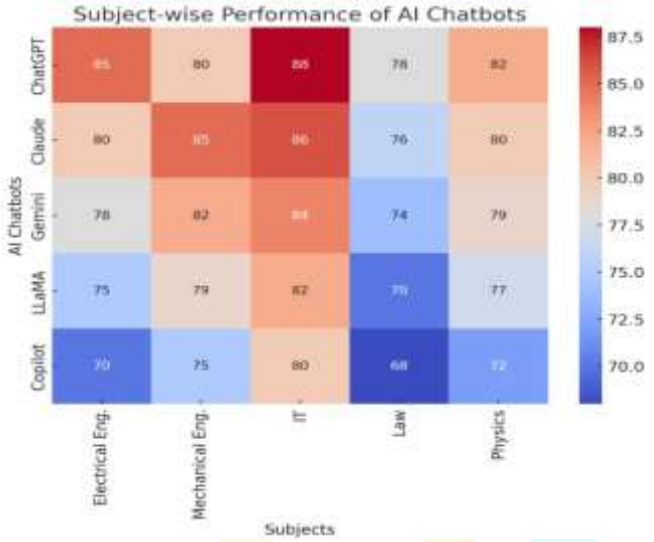
OUTCOMES.

1. Accuracy Trends: Analysis of bar charts revealed that ChatGPT maintained a high accuracy rate across most subjects, whereas Copilot struggled significantly with numerical reasoning tasks.

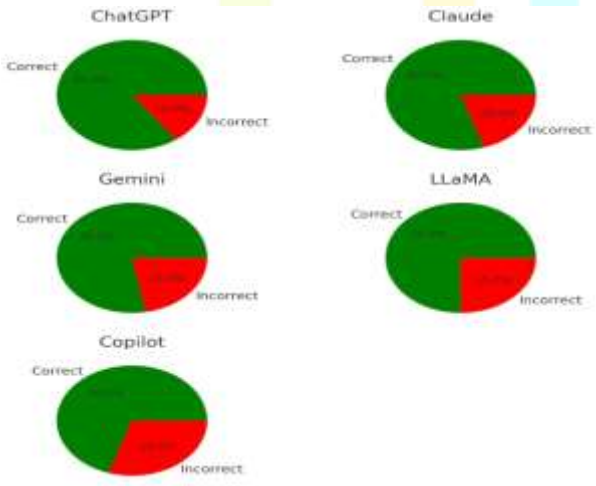
2. Numerical Deviations: Box plot results highlighted that LLaMA and Copilot exhibited the highest numerical discrepancies, whereas Claude and ChatGPT provided more stable calculations.



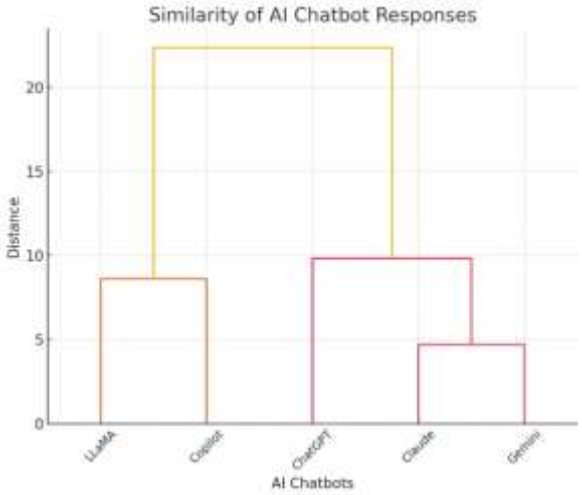
3. Subject-Specific Performance: Heatmap visualizations demonstrated that AI accuracy varied across disciplines, with IT related questions receiving the highest accuracy rates, while mechanical engineering concepts led to the greatest inconsistencies.



4. Correct vs. Incorrect Response Distribution: Pie chart analysis showed the proportion of accurate versus incorrect responses for each chatbot, further revealing inconsistencies in their ability to process domain-specific queries.

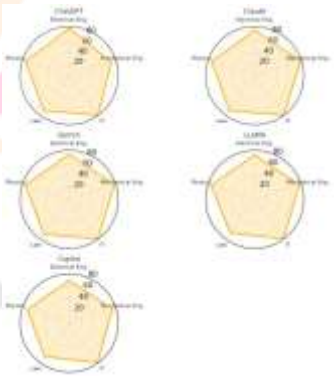


5. Response Similarity Analysis: A dendrogram was used to cluster chatbot responses, indicating that Gemini and Claude provided closer matches in their responses compared to LLaMA and Copilot, which showed greater divergence.



DISCUSSION: LIMITATIONS & FUTURE SCOPE

Despite demonstrating significant potential, AI chatbots still exhibit notable limitations. Numerical problem-solving inconsistencies persist, and legal interpretations often lack the contextual depth required for complex inquiries. Furthermore, AI-generated responses sometimes include multiple interpretations of a question, making accuracy assessments more challenging. Future research should explore AI advancements in subject specific reasoning, leveraging improved training methodologies and domain-specialized fine-tuning to enhance chatbot reliability.



CONCLUSION

This study offers a comparative evaluation of AI chatbots across multiple academic disciplines, emphasizing variations in accuracy and response quality. ChatGPT and Claude demonstrated higher reliability and precision, while LLaMAand Copilot exhibited greater numerical deviations and lower accuracy in certain domains. Understanding these differences is essential for optimizing AI applications in educational and professional settings. As AI technology evolves, future enhancements should aim at improving chatbot adaptability, reasoning capabilities, and contextual depth to better serve users across various fields.

REFERENCES

 AIChatbotsinEducation: A Systematic Literature Review This paper explores the role of AI chatbots in education, discussing their advantages and limitations in academic settings. https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239 023-00426-1
 ASystematicReviewand ComprehensiveAnalysis of AI Chatbot Applications This study presents a comparative analysis of AI chatbots like ChatGPT, Gemini, and Claude, evaluating their usability and accuracy. https://www.mdpi.com/1999-5903/16/7/219
 ComparingandAssessingAI Chatbots' Competence in Economics This research evaluates the accuracy and quality of responses from AI chatbots in the domain of economics and education. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0297804
 TheWho,Why,andHowofAI-BasedChatbotsforLearning and Teaching Asystematic review of AI chatbots' use in higher education, focusing on their effectiveness in learning and knowledge delivery. https://link.springer.com/article/10.1007/s10639-024-13128-6
 AIChatbotsFace-Off: A Comprehensive Comparison This article provides an in-depth comparison of leading AI chatbots, highlighting differences in reasoning, accuracy, and applications. https://ainows.com/ai-chatbots-comparison/