Harnessing AI for Geosciences Education: A Deep Dive into ChatGPT's Impact

Subham Patra[1]*, T Sumit Singha[1], Megh Kanvinde[1], Angana Mazumdar[1] and Swastika Kanjilal[1]

[1] Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India

Abstract

The integration of artificial intelligence language models, particularly ChatGPT, into geosciences education has the potential to transform the learning landscape. This study explores the impact of ChatGPT on geoscience education. The research comprises two phases: first, a survey to understand students' perceptions and usage patterns of ChatGPT, and second, a series of tests to assess its reliability, content generation capabilities, translation abilities, and potential biases.

The survey findings reveal that ChatGPT is gaining popularity among geoscience students, with many using it as a quick information retrieval tool and for content generation tasks. However, students expressed concerns about its accuracy, potential biases, and lack of awareness regarding its limitations. While ChatGPT offers benefits in terms of generating content and streamlining educational tasks, it cannot replace the essential role of human teachers in fostering critical thinking and problem-solving skills. Thus, a balanced approach is crucial. Ethical concerns surrounding ChatGPT include its potential to bypass plagiarism detectors, introduce biases, and raise issues related to data privacy and misinformation. Responsible adoption of AI technologies in education is essential to address these concerns. In conclusion, ChatGPT has the potential to enhance geoscience education, but its implementation should be approached with caution. By understanding its capabilities and limitations, educators can leverage AI technologies to create more engaging, inclusive, and effective learning experiences while upholding academic integrity and ethical standards.

1. Introduction

Artificial intelligence language models have recently witnessed a significant rise in popularity, revolutionizing various domains across multiple sectors (Steenbergen-Hu and Cooper, 2014; Zawacki-Richter et al., 2019; Bengio et al., 2021; Xu et al., 2021; Sallam, 2023). These models have proved their capabilities in learning, judgment, and decision-making, making them invaluable. Prominent examples of the AI language models include BERT (Bidirectional Encoder Representations from Transformers) by Google, T5 (Text-to-Text Transfer Transformer) by Google, and ChatGPT (Generative Pre-trained Transformer) developed by OpenAI. These models are pre-trained on vast datasets from the internet, allowing them to develop a generalized understanding of language and context. The large language models have now set and continue to achieve new benchmarks in natural language processing, empowering computers to process, understand, and generate human-like text.

ChatGPT (Generative Pre-trained Transformer) developed by OpenAI stands out at present as an influential AI language model and has gathered considerable attention since its inception (30th November 2022 – initial release date; https://openai.com). It builds upon the foundation set by its predecessor GPT-3, offering significant improvements in generating contextually relevant and coherent responses resembling natural human dialogue. ChatGPT has had a humongous impact on conversational AI, evident in its enhanced natural language understanding, personalization capabilities, multilingual support, and ability to boost user engagement. OpenAI has made an open-source version of ChatGPT available, allowing developers and researchers to integrate into numerous fields to enhance various processes. In addition, the company has been actively working on the next iteration, GPT-4 which is expected to offer even more sophisticated language understanding and generation capabilities (including image and voice inputs).

The potential applications of ChatGPT in the education sector are vast and hold promising prospects for both students and educators (Zhai, 2022; Sallam, 2023; Kasneci et al., 2023). The chatbot's capabilities are broad and versatile - ranging from tasks like question-answering, language translation, text summarization, etc. (Gilson et al., 2023; Hargreaves, 2023; Jiao et al., 2023) – making it a complete education and research assistant for students. In this study, we aim to investigate the impact of ChatGPT in the geoscience sector. Geoscience education - a specialized field centered on the study of Earth's structure, processes, and history - plays a pivotal role in understanding our planet's past, present, and future. Unlike most educational disciplines, geoscience education presents unique challenges and opportunities due to its reliance on visualizations, hands-on fieldwork, and the need for scientific precision. Given these characteristics of geoscience education, the introduction of AI language models like ChatGPT holds significant promise in terms of data analysis, visual interpretations, and scientific communication. Moreover, ChatGPT's availability at any time allows students to seek help and clarification outside of traditional classroom hours, enhancing their learning experience. In this study, we aim to investigate how ChatGPT's capabilities can/should be harnessed to improve geoscience education through the following objectives:

i. Surveying geoscience students to assess their familiarity with ChatGPT and its features relevant to geoscience education. Additionally, determining their usage frequency,

ii. Testing these features for their accuracy, reliability, and fidelity,

iii. Providing a comprehensive overview of the usability and limitations of ChatGPT in geoscience education.

2. Methods

This ~~entire~~ research was conducted in two phases. The first phase included surveying students to understand their insights of ChatGPT and its applicability and the final phase included testing out these features and comparing them with the general perception. The survey was conducted among

geoscience students in Mumbai, India, focusing on those who were active during and after November 2022 (the initial release date of ChatGPT). Participants belonged to three major institutes in Mumbai that offer courses in geosciences, namely: Indian Institute of Technology, Bombay, St. Xavier's College, and K J Somaiya College of Science and Commerce. Anonymous responses were collected to maintain the authenticity in data. A total of 94 geoscience students took part in the survey, which consisted of 20 questions that aimed to assess their awareness of the model and the frequency of its usage. Open-source software accessible to the surveyed students was primarily utilized in the study to ensure accessibility and reliability.

The reliability of ChatGPT's most used feature – answering questions – was assessed by presenting conceptual and problem-solving questions in geosciences. Additionally, ChatGPT was prompted to attempt questions from the Graduate Aptitude Test in Engineering (GATE) examinations (questions of years 2016, 2018, 2019, 2021). The GATE examination is designed to evaluate a comprehensive understanding of engineering and science for admission into Master's programs in reputed institutes of the country and recruitment by some public sector companies. However, due to the limitations of the free version of ChatGPT, which cannot accept images as prompts, some questions could not be attempted. Moreover, one section that contains aptitude questions (unrelated to geoscience) was excluded.

To evaluate ChatGPT's performance in content generation (infamously used by students to complete assignments requiring mere text generation), the model was asked to generate 200 essays on various genres of geology, such as sedimentology, metamorphic petrology, structural geology, etc. All essays were then assessed for plagiarism using the Grammarly software (Dong and Shi, 2021; http://grammarly.com). Additionally, the essays were tested using GPTZero, a classification model designed to detect whether a document was written by a large language model (http://gptzero.me). GPTZero was trained on diverse human-written and AI-generated text, with a focus on English prose.

107 While GPTZero's accuracy may vary across different use-cases, it has been endorsed as one of the

108 most reliable AI detectors by multiple independent sources, including TechCrunch. Further, repeated

109 analysis (20 times) of the same essay on GPTzero revealed that it is highly precise with its responses,

110 giving the same response every time. It also proved to be efficient in detecting human-generated texts

111 as it successfully recognized them correctly 20 times.

112

113 ChatGPT's writing ability was tested by making it re-write 50 abstracts from published research

114 articles. Text scores provided by Grammarly were used to compare the original and modified articles

115 for linguistic accuracy and quality. A metric 'improved%' was calculated with the formula

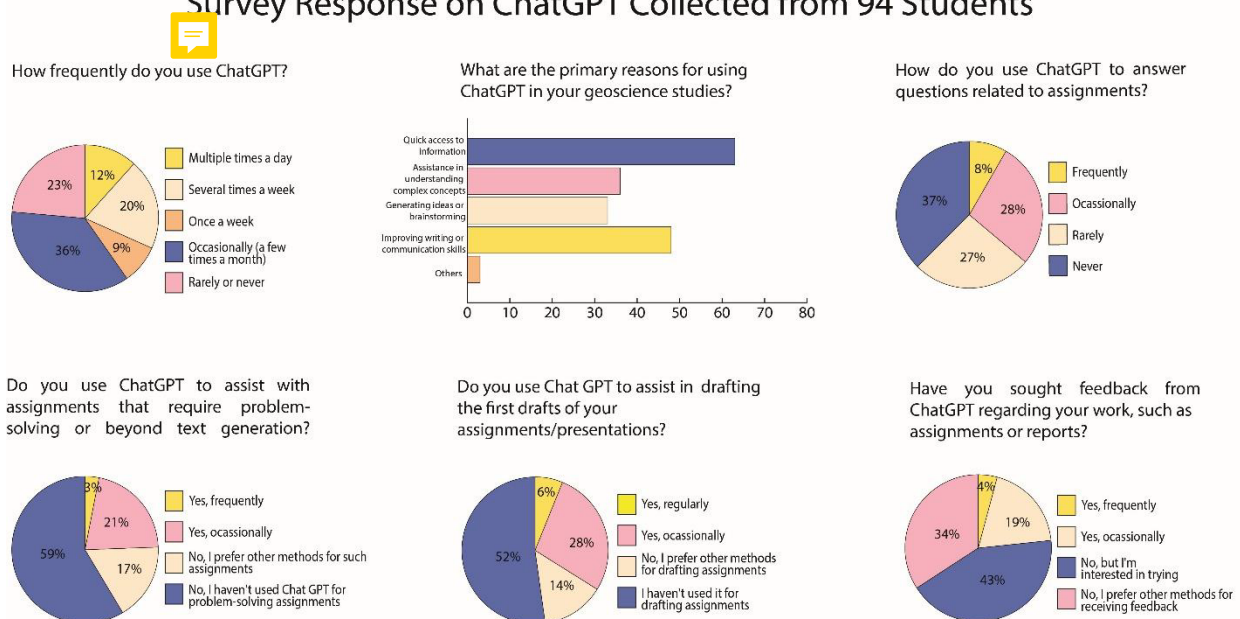
116 "Improved% = (Modified rating - Original rating/Original rating) *100. Furthermore, the model's

117 translation (Jiao et al., 2023) ability was assessed by translating English words, sentences, and

118 paragraphs to Hindi using its inherent features. Google Translate was utilized for comparison

119 purposes. The translated content was reviewed for accuracy by two authors fluent in Hindi. In

120 addition, the model was subjected to bias testing by presenting questions that could have multiple

121 answers, to assess the potential bias in the content it generates.

122

123 Figure 1 Survey Results depicting ChatGPT Utilization in Geosciences among 94 Student Respondents. Many favor frequent use for
124 quick info, but potential for problem-solving remains largely unexplored, highlighting untapped opportunities for its application in
125 academia.

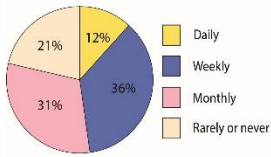
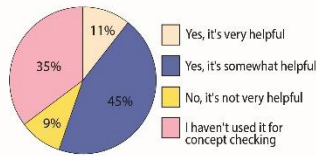
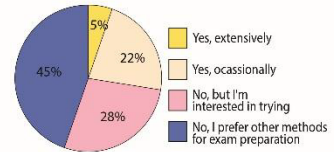
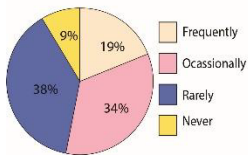
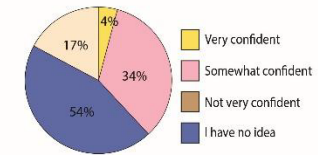



Figure 2 Survey Results depicting ChatGPT Utilization in Geosciences among 94 Student Respondents. In Mumbai, geoscience students frequently use ChatGPT for self-study, but they have reservations about its accuracy, potential data bias, and knowledge limitations when it comes to exam preparation.

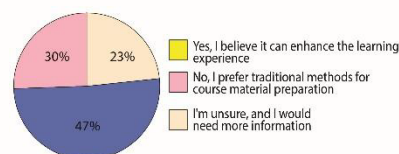
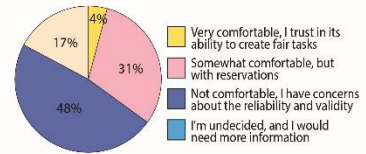
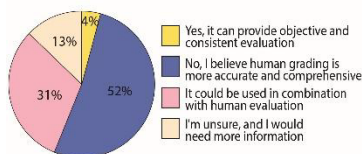
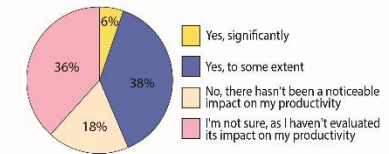



Figure 3 Ethical Perspectives on AI in Academia: Majority of students oppose AI-generated texts evading plagiarism detection, prefer traditional teaching methods, and consider human teachers indispensable, yet appreciate ChatGPT's productivity boost.

3. Results

3.1 Phase 1: Survey Insights of ChatGPT and Its Applicability

A survey was conducted among 94 geoscience students who were active during the release of ChatGPT or after it. The survey aimed to assess the frequency of ChatGPT usage, participants'

awareness of its features, and their perspectives on its potential use for teaching purposes (Fig. 1,2,3; Supplementary file S2).

Findings from the survey revealed that 12% of the students reported using ChatGPT frequently, while approximately 36% were occasional users (Fig. 1). The primary reasons cited for using the AI tool were quick access to information (63) and improvement in writing and communication skills (48). Regarding academic use, around 64% of the students admitted to either not using ChatGPT or using it rarely to answer questions related to assignments (Fig. 1). Furthermore, a significant proportion (59%) of the participants were unaware that ChatGPT could be utilized for tasks beyond simple text generation, such as problem-solving (Fig. 1). An interesting feature of ChatGPT is its capability to generate drafts for assignments, presentations, and talks (Choi et al., 2023). However, 52% of the students reported not using this feature (Fig. 1). Additionally, only 23% of the participants used ChatGPT to receive feedback on their provided documentation (Cotton et al., 2023), with 34% preferring other methods (Fig. 1). During self-study sessions, approximately 48% of the geoscience students in Mumbai reported using ChatGPT at least weekly, and 56% of them found it very useful (Fig. 2). However, around 45% of the students did not use the tool during exam preparation, although 28% expressed interest in using it for this purpose (Fig. 2).

When questioned about the accuracy of the AI bot, 53% of the students encountered instances where ChatGPT produced incorrect results (Fig. 2). Moreover, 54% of the participants believed that the model relied on biased data (Fig. 2). Notably, ChatGPT's knowledge was limited to data up to September 2021, a fact acknowledged by the bot itself, but 61% of the students expressed uncertainty or were unaware of this limitation (Fig. 2). Regarding ethics in academia, 53% of the students opposed the idea of AI-generated texts bypassing plagiarism detectors (Khalil and Er, 2023), while 39% had a neutral opinion on the matter (Fig. 3). ChatGPT's translation feature (Jiao et al., 2023) was utilized by only 31% of the students, with 18% preferring other methods for translation. In terms of teaching

preferences, a majority of students preferred traditional teaching methods (not involving AI) for course material generation (47%), assessment task preparation (48%), and grading (52%) (Supplementary file S1). Additionally, 64% of the participants strongly believed that human teachers were essential for effective learning and that chatbots could not replace them (Fig. 3). Finally, around 44% of the students admitted that ChatGPT had improved their productivity while studying geosciences (Fig. 3).

3.2 Phase 2: Testing ChatGPT Features

3.2.1 Reliability Assessment of ChatGPT's Question-Answering Feature

ChatGPT's capability to function as a search engine and explain conceptual questions in geology was tested to assess its accuracy and usefulness for self-study by students. The results of these exercises revealed both strengths and limitations in this feature.

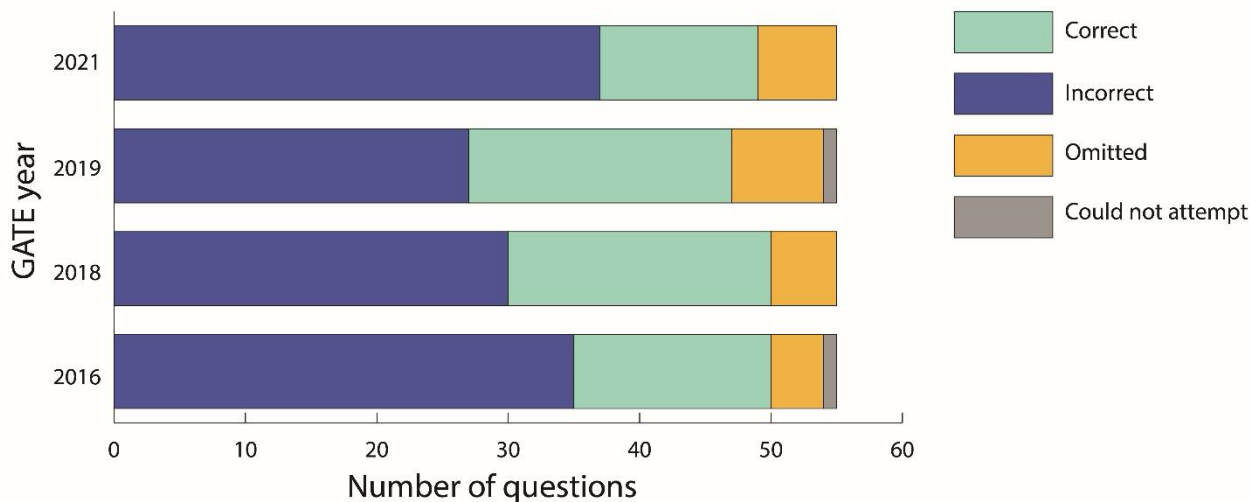
When asked basic conceptual questions on geology, ChatGPT provided correct and well-structured explanations, demonstrating its effectiveness as a self-study tool for students. Its ability to explain complex concepts in a clear manner can be valuable for enhancing students' understanding. A critical limitation observed during the exercises was ChatGPT's inability to generate images. In geology, where visual representations are often essential for comprehension, this limitation hinders the model's effectiveness in providing a comprehensive learning experience. When asked to generate the geological time scale, ChatGPT displayed high inaccuracy, mislabeling time units, and omitting important information in many instances (Supplementary file S2). This inaccuracy raises concerns regarding the reliability of the information provided for important geological concepts. ChatGPT was tasked with generating references on specific geological topics. The results showed mixed accuracy, with some references being incorrect and fake. For instance, when asked to provide references on

190   end-Cretaceous stress environments, three out of five references were wrong and not genuine

191   (Supplementary file S2). Similarly, for scientific articles discussing the role of carbon isotopes in

192   interpreting the 'big five' mass extinctions, five out of ten references were incorrect (Supplementary

193   file S2).

Number of yearwise GATE questions attempted correctly and incorrectly by ChatGPT

*Questions marked as "omitted" have not been included as they contain diagrams required to answer them. The version of ChatGPT used in this study does not accept multimedia in the questions, nor does it include multimedia in its responses. The diagrams mentioned here are either in the question or the given answer choices.

*The questions marked as "could not attempt" are the ones where ChatGPT explicitly stated that it could not answer the question because of insufficient data.

194

195 Figure 4 ChatGPT's performance in GATE Questions: Stacked bar graph illustrating correct and incorrect answers, highlighting its
196 struggles in problem-Solving with a 20.4% accuracy rate.

197   ChatGPT's performance in solving GATE examination questions was evaluated, and it scored poorly

198   with an average of 36.44% (Highest score – 41.6%, 2018; lowest score- 29.4%, 2021). The model

199   struggled particularly with questions that required problem-solving, such as numericals, achieving

200   only a 20.4% accuracy rate (Fig. 4; Supplementary file S3).

201

202   **3.2.2 Content Generation Performance Evaluation**

203

204   The content generation feature of ChatGPT emerged as the most utilized by geoscience students,

205   especially for generating content related to assignments, scripts, and during self-study sessions. To

206   evaluate the performance of this feature, two exercises were conducted, each focusing on different

207   aspects of content generation.

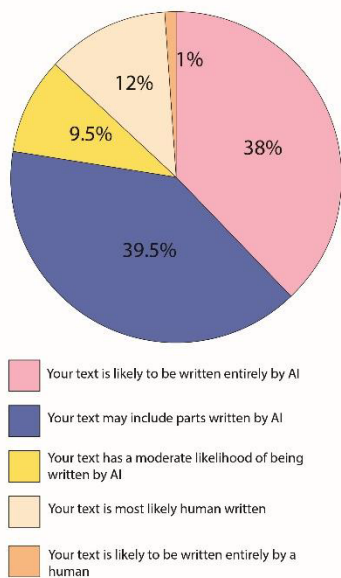



Figure 5   Assessing ChatGPT's Content Generation: A pie chart showcasing GPT Zero responses reveals the accuracy in detecting AI-generated content. Furthermore, a histogram of plagiarism scores for the same essays illustrates the low level of plagiarism in ChatGPT's content.

ChatGPT was prompted to generate 200 essays covering various topics across different domains of geosciences (Fig. 5; Supplementary file S4). The essays were subsequently analyzed for plagiarism using Grammarly's built-in features. The results showed an average plagiarism rate of 1.46%, indicating a low level of plagiarised content in the generated essays. Most essays had minimal or no plagiarism, with 51% having less than 1% copied content and 94.5% having less than 4% copied content. To further evaluate the authenticity of the generated content, GPTZero, a classification model for detecting AI-generated text, was employed. According to GPTZero's analysis, approximately 78% of the essays were successfully identified as either entirely generated by AI or containing AI-generated portions. However, interestingly, around 12% of the essays were identified as mostly written by humans, showcasing the model's capability to produce human-like content. It is worth noting that only 1% of the essays were mis-detected as entirely written by humans.

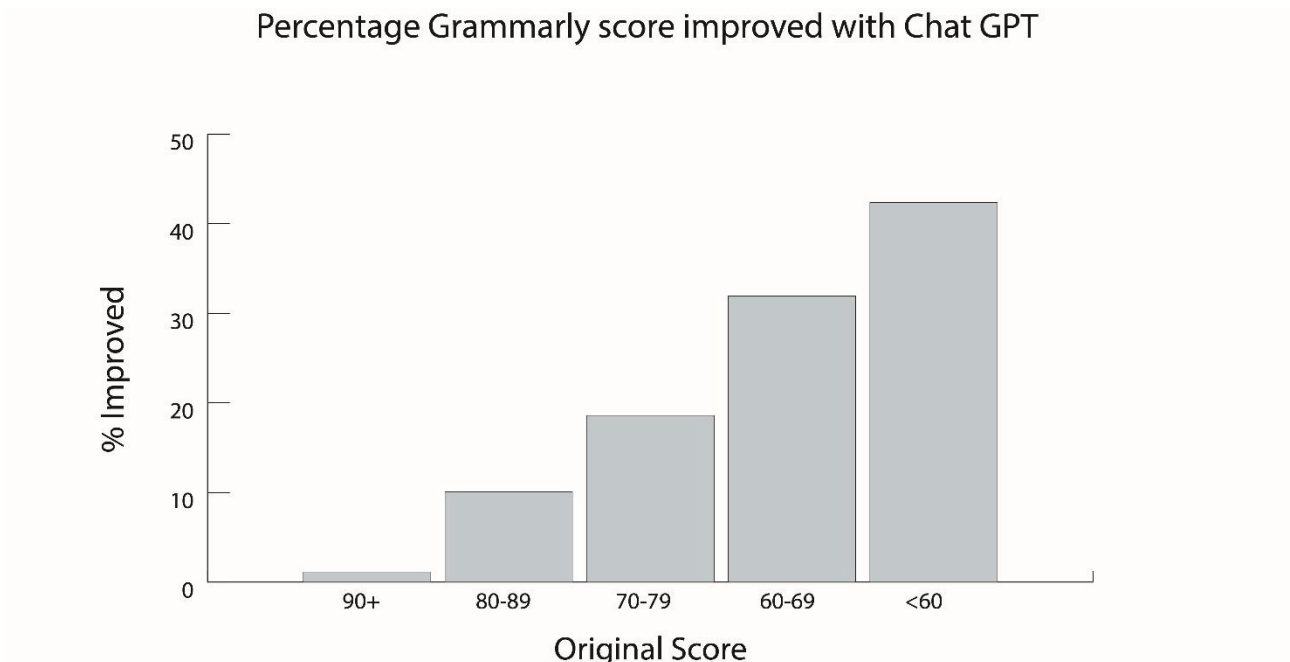
Published abstracts (100) were modified using ChatGPT, and their Grammarly Scores were compared before and after the modifications (Fig. 6; Supplementary file S4). The results indicated that most abstracts (91%) showed an improvement in their Grammarly Scores after being modified by

227 ChatGPT. The average improvement observed throughout the abstracts was 16.21%. An intriguing

228 observation was that ChatGPT significantly improved the writing of poorly written texts (with low

229 initial Grammarly Scores), thus following an exponential curve for improvement. All the abstracts

230 that would not get improved later or showed minor improvements, originally had a score of more

231 than 80, suggesting that the model is more effective in enhancing poorly written texts.

232

233 Figure 6 ChatGPT's impact on content quality enhancement. Analysis of Grammarly Scores before and after ChatGPT modifications
234 reveals a 16.21% average improvement, with notable effectiveness in enhancing poorly written texts, as demonstrated by a significant
235 improvement in previously low-scoring abstracts.

236

237 ### 3.2.3 Translating Ability Assessment

238

239 To evaluate ChatGPT's translating service, we tested its ability to translate geological words and

240 sentences from English to Hindi, which is commonly spoken in Mumbai and India. The translations

241 were assessed using the metrics 'Accurate,' 'Moderate,' and 'Poor' to comment on the quality of the

242 translations. Out of the geological words translated, 54% of them were accurately translated, meaning

243 the Hindi translations were correct and aligned with their intended meanings (Supplementary file S5).

244 However, a notable concern was that 26% of the translations were categorized as 'Poor,' indicating

245 incorrect translations. Moreover, 20% of the translations were classified as 'Moderate,' implying that

although the translations were somewhat correct, they were not entirely accurate and might have slightly deviated from their intended meanings. For the translation of English sentences related to geological terms, only 60% of the sentences were 'Accurately' translated, where the Hindi translations correctly conveyed the intended meanings of the sentences (Supplementary file S5). A concerning observation was made in 38% of the sentences, where ChatGPT did not translate critical terms and instead used them as they were, in English. This failure to translate crucial terms hinders the overall effectiveness of the translated sentences.

3.2.4 Bias Testing

ChatGPT's training process involves learning from a vast range of internet text, including articles, books, and websites, capturing both factual information and subjective perspectives available online. As with any AI language model, the training data can potentially include biased language or reflect existing biases present in society.

To assess ChatGPT's response accuracy and potential biases in the context of geosciences, two exercises were conducted. These exercises highlight the importance of understanding potential biases and limitations in AI language models like ChatGPT when dealing with subject areas that can have diverse perspectives and interpretations. In the first exercise, the model was asked to generate ten references of scientific articles discussing the role of carbon isotopes in interpreting the 'big five' mass extinctions (Supplementary file S2). However, the response exhibited some bias, as six out of the ten references focused solely on the Permian-Triassic mass extinction. An unbiased response should have contained references from articles discussing at least one of each of the 'big five' mass extinctions, providing a more balanced representation.

The second exercise involved asking ChatGPT about the cause of the Cretaceous-Paleogene boundary mass extinction, a topic with two competing schools of thought (Supplementary file S2). One group supports an asteroid impact as the cause (Schulte et al., 2010), while the other advocates for the Deccan volcanism hypothesis (Keller et al., 2020). The model predominantly discussed the asteroid impact and its repercussions as the primary cause of the mass extinction in five out of six short paragraphs. Only in the end, it briefly mentioned volcanic activity and long-term environmental changes as contributing factors. An unbiased response would have evenly presented both possible causes and perhaps included a note about the prevailing opinion regarding the event's cause.

4. Discussion

4.1 Benefits and limitations of ChatGPT in geoscience education

In recent years, the development of large language models, including the widely used ChatGPT, has revolutionized various domains, including education (Farrokhnia et al., 2023; Lo, 2023; Elbanna and Armstrong, 2023; Li et al., 2023). These transformer-based models have been pre-trained on massive datasets of text, enabling them to generate human-like text, answer questions, and assist with translation and summarization (Lo, 2023). In the field of geosciences, where understanding complex processes and historical events requires significant imagination and critical thinking, such models hold great potential to play a vital role in education. However, it is essential to examine their capabilities and limitations to ensure their effective use in geoscience education. Our study focused on understanding geoscience students' perceptions and usage patterns of ChatGPT in Mumbai, India (Fig. 1,2,3). The results revealed that approximately 32% of geoscience students admitted to using the chatbot several times a week, indicating its growing popularity among students in a metropolitan city like Mumbai.

The majority of students found the chatbot useful as a 'search engine' to quickly access information, outperforming traditional methods like Google Search due to its interactive nature and concise responses (Fig. 1). Nevertheless, our findings highlighted several limitations that warrant careful consideration. ChatGPT's responses lacked proper scientific references, and inaccuracies were observed, with instances of the model generating non-existent article references and bibliographic details (Sect. 3.2.1). Such concerns have been previously reported in the literature, indicating the need for caution when relying on ChatGPT for academic tasks in specialized domains like geosciences. Additionally, ChatGPT's performance in solving GATE examination questions was found to be moderate, particularly struggling with numerical-based questions, with only 20.4% accuracy (Fig. 4). This emphasizes the importance of cross-referencing and validating information from alternative sources when dealing with critical assessments and evaluations.

On a positive note, ChatGPT exhibited excellence in content generation and language editing (Fig. 5,6). The model generated well-written texts with improved Grammarly scores, showcasing its potential as a valuable tool for enhancing students' writing and communication skills. Moreover, its translating ability equaled traditional services like Google's, given its human-like communication capabilities (Supplementary S5).

However, an important aspect that demands attention is biases in ChatGPT's responses (Tlili et al., 2023; Baidoo-Anu and Owusu Ansah, 2023). The model's reliance on a large corpus of data can lead to biased outcomes, with responses disproportionately focused on specific contents, such as the Permian-Triassic mass extinction and the impact as the cause of the fifth mass extinction in our examples (Sect. 3.2.4). This bias could stem from the prevalence of certain topics in the training data, possibly influenced by the availability of published literature and media coverage. , raising concerns about the reliability of responses on certain topics. Future research could quantify data sources to better understand and address bias in AI language models like ChatGPT.

Over-reliance on AI, including ChatGPT, may hinder the development of essential skills like critical thinking, problem-solving, imagination, and research abilities in students. Worryingly, a considerable percentage of students were unaware of the possibility of biased (46%), incorrect (47%), and outdated (62%) responses from ChatGPT, highlighting the need for educational institutes to conduct awareness sessions (Fig. 2). Promoting responsible usage and critical evaluation of AI language models will help students harness the benefits while being mindful of the limitations.

4.2 Pedagogical considerations

The introduction of large language models like ChatGPT has ushered in a new era of technological advancement in education. As technology continues to evolve rapidly, it inevitably impacts education systems worldwide, prompting educators to explore the implications of incorporating AI technologies into teaching and learning processes (Ausat, 2023). ChatGPT, as a powerful artificial intelligence system capable of processing and generating sophisticated text, has the potential to revolutionize the traditional classroom dynamic and raise critical questions about the role of teachers in the learning process (Ausat et al., 2023; Fauzi et al., 2023).

Teachers play multifaceted roles beyond being instructors, serving as mentors and role models for students (Zen et al., 2023). The introduction of ChatGPT and other AI technologies into the geosciences educational landscape has the potential to complement and augment these roles in various ways. One of the notable contributions of ChatGPT lies in its ability to provide high-quality reading materials tailored to students' comprehension levels (Kasneci et al., 2023). By processing natural language, ChatGPT can produce texts that are easy to understand, making it a valuable tool for teachers in creating customized learning experiences. 23% of the students did show interest in enhancing their learning experience through this human-AI collaboration (Fig. 3). Additionally, the

AI model can automatically generate questions and tests that match students' proficiency levels, streamlining the assessment process (Cooper, 2023; Tlili et al., 2023). This can save teachers time and effort while providing relevant and differentiated assessments for students. ChatGPT also holds promise in supporting research and writing tasks. It can aid teachers in identifying and correcting errors, highlighting grammatical inconsistencies, and suggesting personalized improvement strategies. Furthermore, the AI model can generate summaries and outlines of complex texts, assisting educators in emphasizing key points for further exploration and understanding. It can also be instrumental in identifying areas where students are struggling, facilitating targeted instruction for their improvement. However, a large proportion of students (53%) currently doubt the validity and reliability of AI generated assessments (Fig. 2).

While ChatGPT's potential to streamline various educational tasks is evident, it is important to recognize its limitations. The AI model can only generate text-based responses and lacks the ability to provide live explanations or real-time examples, which are inherent to human teachers' interactions with students (Herft, 2023). Consequently, ChatGPT's usage should be seen as an adjunct to, rather than a replacement for, the vital role teachers play in fostering critical thinking, problem-solving, and creativity in students. Thankfully, most of the students (~64%) believe that human teachers are essential for effective learning (Fig. 3).

As educators embrace the integration of technology in the classroom, they must be proactive in upskilling their competencies and practices to effectively leverage AI's benefits (Haleem et al., 2022). ChatGPT, as a powerful tool, necessitates thoughtful design strategies to balance human and machine intelligence in collaborative learning environments. This demands investigation into how teachers can effectively work together with large language models to achieve desired learning objectives (Salas-Pilco et al., 2022). Furthermore, educators need to explore innovative ways of using ChatGPT and other AI technologies to promote personalized learning experiences (Hwang & Chang, 2021). By

using AI-generated adaptive feedback and course materials, teachers can better cater to students' individual needs and learning preferences. Moreover, they can use large language models to create targeted practice problems and quizzes, ensuring students achieve mastery in the subject matter. As the adoption of AI in education continues to evolve, future research should focus on understanding the potential of large language models in supporting teaching practices. Investigating different human-machine collaboration strategies will be crucial in harnessing the benefits of AI while preserving the essential human touch in the teaching-learning process. The aim should be to strike a balance between AI assistance and human interaction, resulting in more engaging, inclusive, and effective learning experiences for students.

4.3 Ethical and societal implications

The integration of AI, particularly generative AI like ChatGPT, into educational settings raises numerous ethical concerns that have garnered attention from international organizations and researchers (Tlili et al., 2023; Lo et al., 2023). Among the critical concerns identified, one issue stands out prominently - the potential for AI-generated texts to bypass plagiarism detectors, an alarming fact supported by our research, which indicates that around 50% of students do not support the notion that AI-generated texts can circumvent plagiarism detection measures (Fig. 3). This phenomenon poses a significant threat to academic integrity and the fundamental purpose of assessment, which is to evaluate students' original work and knowledge accurately. The implications of AI-generated content being undetectable by plagiarism detection applications (e.g., Turnitin and iThenticate) have serious consequences, as students using ChatGPT can obtain an unfair advantage over their peers who put in genuine efforts to produce original work (Bašíc et al., 2023; Cotton et al., 2023). Furthermore, instructors find it challenging to evaluate and follow up on students' learning progress when AI-generated content is involved, potentially undermining the overall effectiveness of the educational system.

400

401   Beyond the issue of plagiarism, the implementation of AI in education also brings to light concerns

402   about bias and inequalities. AI-assisted chatbots like ChatGPT can inadvertently perpetuate biases

403   present in the training data, leading to the reinforcement of existing inequalities in education (Zhai,

404   2022). This raises ethical questions about ensuring fairness and equal opportunities for all students

405   and highlights the need for developing AI systems that are free from inherent biases and

406   discrimination. Another significant ethical consideration is the privacy and security of students' data.

407   AI technologies collect and process vast amounts of data from users, including students, to improve

408   their performance. However, there are valid concerns about how this data is utilized, stored, and

409   protected. Safeguarding students' privacy and ensuring the secure handling of their data is of utmost

410   importance to maintain trust in AI technologies in educational environments.

411

412   Furthermore, the potential for AI-generated content to contain errors or even fake information raises

413   ethical questions regarding the dissemination of misinformation in scientific publications and

414   academic work (Tlili et al., 2023; Liebrenz et al., 2023). While ChatGPT's responses are not exact

415   copies of specific texts, their similarity to existing sources can lead to misleading content. This

416   emphasizes the necessity of developing ethical guidelines for the use of AI in education to promote

417   accuracy and credibility in academic work.

418

419   Despite these concerns, it is essential to acknowledge the positive aspects of integrating AI in

420   education. ChatGPT and similar AI technologies have the potential to enhance instruction delivery

421   and learning practices, benefitting both teachers and students in various educational tasks, such as

422   preparing teaching materials, creating quizzes, and offering personalized learning experiences

423   (Kasneci et al., 2023). However, balancing the advantages with the ethical challenges is essential.

424   The decision by New York City to ban ChatGPT in schools due to concerns about cheating in

425   homework and assignments highlights the need for careful consideration and responsible use of AI

technologies in educational contexts (The Guardian, 2023). Rather than outright bans, engaging in informed discussions and collaborating with experts from different fields, including education, security, and psychology, is crucial to fostering a deeper understanding of AI's implications and responsible adoption of chatbots like ChatGPT.

5. Conclusions

The integration of artificial intelligence language models like ChatGPT into geosciences education presents both opportunities and challenges. Our study aimed to explore the impact of ChatGPT on geoscience education, particularly among students in Mumbai, India. The findings revealed that ChatGPT is gaining popularity among geoscience students, with many utilizing it as a search engine for quick access to information and for content generation tasks.

However, the study also highlighted several limitations and ethical concerns that need to be addressed. ChatGPT's responses lacked proper scientific references, and inaccuracies were observed in some instances, raising concerns about the reliability of the information provided. Additionally, biases in the model's responses were evident, which can have implications for academic integrity and the reinforcement of existing inequalities in education.

Pedagogically, ChatGPT can be a valuable tool for educators to provide customized learning experiences and streamline various educational tasks. However, it is crucial to recognize that AI cannot replace the vital role of human teachers in fostering critical thinking, problem-solving, and creativity in students.

Ethically, there is a need for guidelines to address concerns about plagiarism, bias, data privacy, and the dissemination of misinformation. Responsible use of AI technologies in education should be

452 promoted, and educators must be proactive in upskilling to effectively leverage AI's benefits while

453 mitigating its limitations.

454

455 In conclusion, ChatGPT holds promise in enhancing geosciences education, but its implementation

456 should be done thoughtfully and responsibly. By understanding its capabilities and limitations,

457 educators can leverage AI technologies to create more engaging, inclusive, and effective learning

458 experiences for students while maintaining academic integrity and ethical standards.

459 **Author Contribution**

460

Contributor role	Contributors
Conceptualization	Subham Patra, T Sumit Singha, Megh Kanvinde
Data curation	Subham Patra, T Sumit Singha, Megh Kanvinde, Swastika Kanjilal
Formal analysis	Subham Patra, Megh Kanvinde
Funding acquisition	Subham Patra
Investigation	Subham Patra
Methodology	Subham Patra, Megh Kanvinde, T Sumit Singha, Swastika Kanjilal
Project administration	Subham Patra

Resources	Subham Patra
Supervision	Subham Patra
Validation	Subham Patra
Visualization	Subham Patra, Angana Mazumder
Writing – original draft preparation	Subham Patra
Writing – review & editing	Subham Patra, Angana Mazumder, T Sumit Singha, Megh Kanvinde

461
462

Competing interests:

The authors declare that they have no known competing financial interests or personal relationships

that could have appeared to influence the work reported in this paper.

466

Ethical statements:

The data used in this study are public and have minimal risk to the individual users from this research.

All students who participated in the survey were provided with participant information. Student

identities have been anonymised throughout the study.

471

Acknowledgements:

We sincerely thank the two anonymous reviewers for their constructive critique. Authors acknowledge the Department of Earth Sciences, IIT Bombay for letting us use assets necessary for carrying out this study. Authors also thank all the anonymous students who participated in the survey for this project.

References:

Xu, L., Sanders, L., Li, K. and Chow, J.C., 2021. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR cancer, 7(4), p.e27850. DOI: 10.2196/27850

Zawacki-Richter, O., Marín, V.I., Bond, M. and Gouverneur, F., 2019. Systematic review of research on artificial intelligence applications in higher education–where are the educators?. International Journal of Educational Technology in Higher Education, 16(1), pp.1-27. DOI: 10.1186/s41239-019-0171-0

Bengio, Y., Lecun, Y. and Hinton, G., 2021. Deep learning for AI. Communications of the ACM, 64(7), pp.58-65.

Steenbergen-Hu, S. and Cooper, H., 2014. A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. Journal of educational psychology, 106(2), p.331. DOI: 10.1037/a0034752

Zhai, X., 2022. ChatGPT user experience: Implications for education. Available at SSRN 4312418. DOI: 10.2139/ssrn.4312418

493   Sallam, M., 2023. The utility of ChatGPT as an example of large language models in healthcare

494   education, research and practice: Systematic review on the future perspectives and potential

495   limitations. medRxiv, pp.2023-02. DOI: 10.1101/2023.02.19.23286155

496   Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh,

497   G., Günnemann, S., Hüllermeier, E. and Krusche, S., 2023. ChatGPT for good? On opportunities and

498   challenges of large language models for education. Learning and Individual Differences, 103,

499   p.102274. DOI: 10.1016/j.lindif.2023.102274

500   Gilson, A., Safranek, C.W., Huang, T., Socrates, V., Chi, L., Taylor, R.A. and Chartash, D., 2023.

501   How does ChatGPT perform on the United States medical licensing examination? The implications

502   of large language models for medical education and knowledge assessment. JMIR Medical

503   Education, 9(1), p.e45312. DOI: 10.2196/45312

504   Hargreaves, S., 2023. 'Words Are Flowing Out Like Endless Rain Into a Paper Cup': ChatGPT &

505   Law School Assessments. The Chinese University of Hong Kong Faculty of Law Research Paper,

506   (2023-03). DOI: 10.2139/ssrn.4359407

507   Jiao, W., Wang, W., Huang, J.T., Wang, X. and Tu, Z.P., 2023. Is ChatGPT a good translator? Yes

508   with GPT-4 as the engine. arXiv preprint arXiv:2301.08745. DOI: 10.48550/arXiv.2301.08745

509   Dong, Y. and Shi, L., 2021. Using Grammarly to support students' source-based writing

510   practices. Assessing Writing, 50, p.100564. DOI: 10.1016/j.asw.2021.100564

511   Choi, J.H., Hickman, K.E., Monahan, A. and Schwarcz, D., 2023. Chatgpt goes to law

512   school. Available at SSRN. DOI: 10.2139/ssrn.4335905

513   Cotton, D.R., Cotton, P.A. and Shipway, J.R., 2023. Chatting and cheating: Ensuring academic

514   integrity in the era of ChatGPT. Innovations in Education and Teaching International, pp.1-12. DOI:

515   10.1080/14703297.2023.2190148

516 Khalil, M. and Er, E., 2023. Will ChatGPT get you caught? Rethinking of plagiarism detection. arXiv

517 preprint arXiv:2302.04335. DOI: 10.48550/arXiv.2302.04335

518 Jiao, W., Wang, W., Huang, J.T., Wang, X. and Tu, Z., 2023. Is ChatGPT a good translator? A

519 preliminary study. arXiv preprint arXiv:2301.08745. DOI: 10.48550/arXiv.2301.08745

520 Schulte, P., Alegret, L., Arenillas, I., Arz, J.A., Barton, P.J., Bown, P.R., Bralower, T.J., Christeson,

521 G.L., Claeys, P., Cockell, C.S. and Collins, G.S., 2010. The Chicxulub asteroid impact and mass

522 extinction at the Cretaceous-Paleogene boundary. Science, 327(5970), pp.1214-1218. DOI:

523 10.1126/science.1177265

524 Keller, G., Mateo, P., Monkenbusch, J., Thibault, N., Punekar, J., Spangenberg, J.E., Abramovich,

525 S., Ashckenazi-Polivoda, S., Schoene, B., Eddy, M.P. and Samperton, K.M., 2020. Mercury linked

526 to Deccan Traps volcanism, climate change and the end-Cretaceous mass extinction. Global and

527 Planetary Change, 194, p.103312. DOI: 10.1016/j.gloplacha.2020.103312

528 Farrokhnia, M., Banihashem, S.K., Noroozi, O. and Wals, A., 2023. A SWOT analysis of ChatGPT:

529 Implications for educational practice and research. Innovations in Education and Teaching

530 International, pp.1-15. DOI: 10.1080/14703297.2023.2195846 CrossMark

531 Lo, C.K., 2023. What is the impact of ChatGPT on education? A rapid review of the literature.

532 Education Sciences, 13(4), p.410. DOI: 10.3390/educsci13040410

533 Elbanna, S. and Armstrong, L., 2023. Exploring the integration of ChatGPT in education: adapting

534 for the future. Management & Sustainability: An Arab Review. DOI: 10.1108/MSAR-03-2023-0016

535 Li, L., Ma, Z., Fan, L., Lee, S., Yu, H. and Hemphill, L., 2023. ChatGPT in education: A discourse

536 analysis of worries and concerns on social media. arXiv preprint arXiv:2305.02201. DOI:

537 10.48550/arXiv.2305.02201

538   Tlili, A., Shehata, B., Adarkwah, M.A., Bozkurt, A., Hickey, D.T., Huang, R. and Agyemang, B.,

539   2023. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education.

540   Smart Learning Environments, 10(1), p.15. DOI: 10.1186/s40561-023-00237-x

541   Baidoo-Anu, D. and Owusu Ansah, L., 2023. Education in the era of generative artificial intelligence

542   (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available

543   at SSRN 4337484.

544   Ausat, A.M.A., Massang, B., Efendi, M., Nofirman, N. and Riady, Y., 2023. Can chat GPT replace

545   the role of the teacher in the classroom: A fundamental analysis. Journal on Education, 5(4),

546   pp.16100-16106. DOI: 10.31004/joe.v5i4.2745

547   Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A.M.A. and Hatta, H.R., 2023. Analysing the role of

548   ChatGPT in improving student productivity in higher education. Journal on Education, 5(4),

549   pp.14886-14891. DOI: 10.31004/joe.v5i4.2563

550   Zen, A., Kusumastuti, R., Metris, D., Gadzali, S.S. and Ausat, A.M.A., 2023. Implications of

551   Entrepreneurship Education as a Field of Study for Advancing Research and Practice. Journal on

552   Education, 5(4), pp.11441-11453. DOI: 10.31004/joe.v5i4.2091

553   Cooper, G., 2023. Examining science education in chatgpt: An exploratory study of generative

554   artificial intelligence. Journal of Science Education and Technology, 32(3), pp.444-452. DOI:

555   10.1007/s10956-023-10039-y

556   Herft, A., 2023. A Teacher's Prompt Guide to ChatGPT aligned with'What Works Best'.

557   Haleem, A., Javaid, M., Qadri, M.A. and Suman, R., 2022. Understanding the role of digital

558   technologies in education: A review. Sustainable Operations and Computers, 3, pp.275-285. DOI:

559   10.1016/j.susoc.2022.05.004

560   Salas-Pilco, S.Z., Xiao, K. and Hu, X., 2022. Artificial intelligence and learning analytics in teacher

561   education: A systematic review. Education Sciences, 12(8), p.569. DOI: 10.3390/educsci12080569

562 Hwang, G.J. and Chang, C.Y., 2021. A review of opportunities and challenges of chatbots in

563 education. Interactive Learning Environments, pp.1-14. DOI: 10.1080/10494820.2021.1952615

564 Bašić, Ž., Banovac, A., Kružić, I. and Jerković, I., 2023. Better by You, better than Me? ChatGPT-3

565 as writing assistance in students' essays. DOI: 10.48550/arXiv.2302.04536

566 Liebrenz, M., Schleifer, R., Buadze, A., Bhugra, D. and Smith, A., 2023. Generating scholarly content

567 with ChatGPT: ethical challenges for medical publishing. The Lancet Digital Health, 5(3), pp.e105-

568 e106. DOI: 10.1016/S2589-7500(23)00019-5

569 Yang, M., 2023. New York City schools ban AI chatbot that writes essays and answers prompts. The

570 Guardian, 6.