https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

Geoscience Communication Discussions
Open Access
EGU

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

*Corresponding author: Subham Patra (subham96@iitb.ac.in)

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.

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

Geoscience Communication Discussions Open Access EGU

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 pretrained 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 opensource 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

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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 (Supplementary file S2). 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. While GPTZero's accuracy may vary across different use-cases, it has been endorsed as one of the most reliable AI detectors by multiple independent sources, including TechCrunch. Further, repeated analysis (20 times) of the same essay on GPTzero revealed that it is highly precise with its responses,

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

giving the same response every time. It also proved to be efficient in detecting human-generated texts as it successfully recognized them correctly 20 times.

ChatGPT's writing ability was tested by making it re-write 50 abstracts from published research articles. Text scores provided by Grammarly were used to compare the original and modified articles for linguistic accuracy and quality. A metric 'improved%' was calculated with the formula "Improved% = (Modified rating - Original rating/Original rating) *100. Furthermore, the model's translation (Jiao et al., 2023) ability was assessed by translating English words, sentences, and paragraphs to Hindi using its inherent features. Google Translate was utilized for comparison purposes. The translated content was reviewed for accuracy by two authors fluent in Hindi. In addition, the model was subjected to bias testing by presenting questions that could have multiple answers, to assess the potential bias in the content it generates.

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).

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

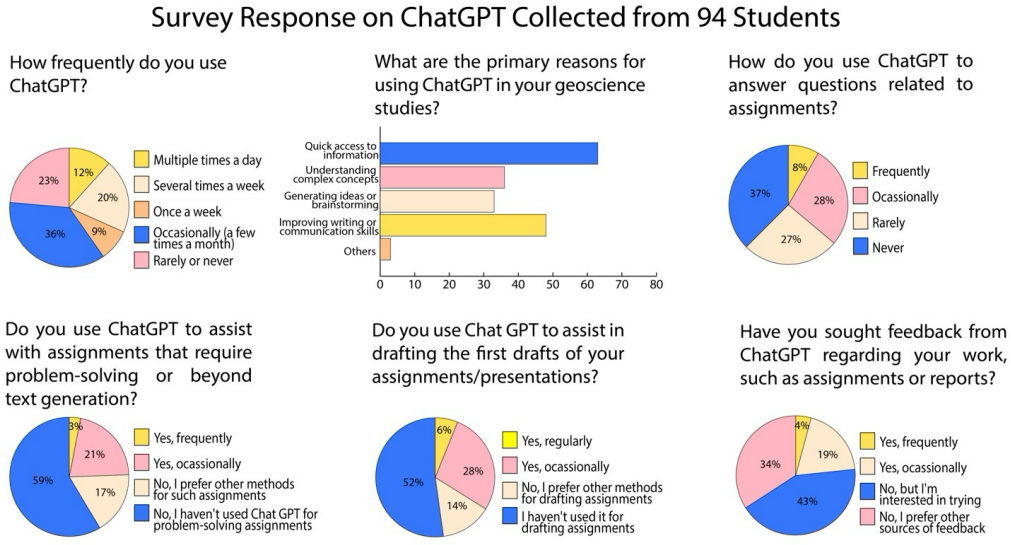


Figure 1 Survey Results depicting ChatGPT Utilization in Geosciences among 94 Student Respondents. Many favour frequent use for quick info, but potential for problem-solving remains largely unexplored, highlighting untapped opportunities for its application in academia.

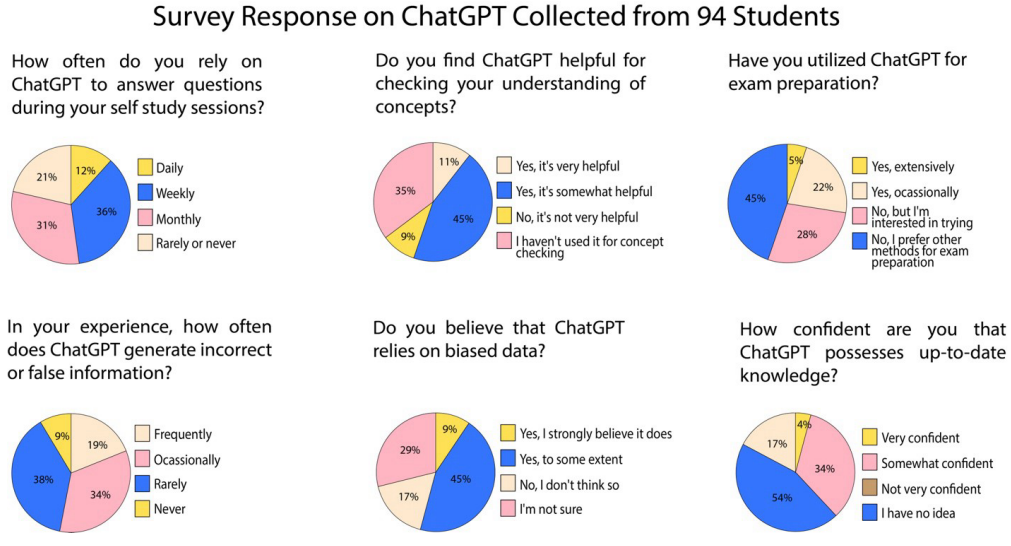


Figure 2 Survey Results depicting ChatGPT Utilization in Geosciences among 94 Student Respondents. In Mumbai, geoscience students frequently use ChatGPT for self-study, but they have reservations about its accuracy, potential data bias, and knowledge limitations when it comes to exam preparation.

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

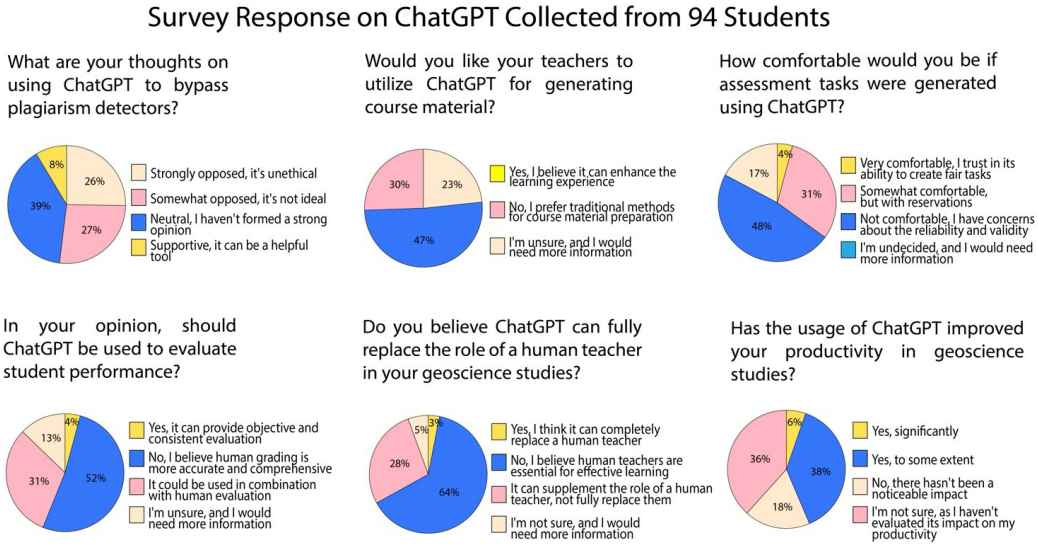


Figure 3 Ethical Perspectives on AI in Academia: Majority of students oppose AI-generated texts evading plagiarism detection, prefer traditional teaching methods, and consider human teachers indispensable, yet appreciate ChatGPT's productivity boost.

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).

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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 end-Cretaceous stress environments, three out of five references were wrong and not genuine (Supplementary file S2). Similarly, for scientific articles discussing the role of carbon isotopes in interpreting the 'big five' mass extinctions, five out of ten references were incorrect (Supplementary file S2).

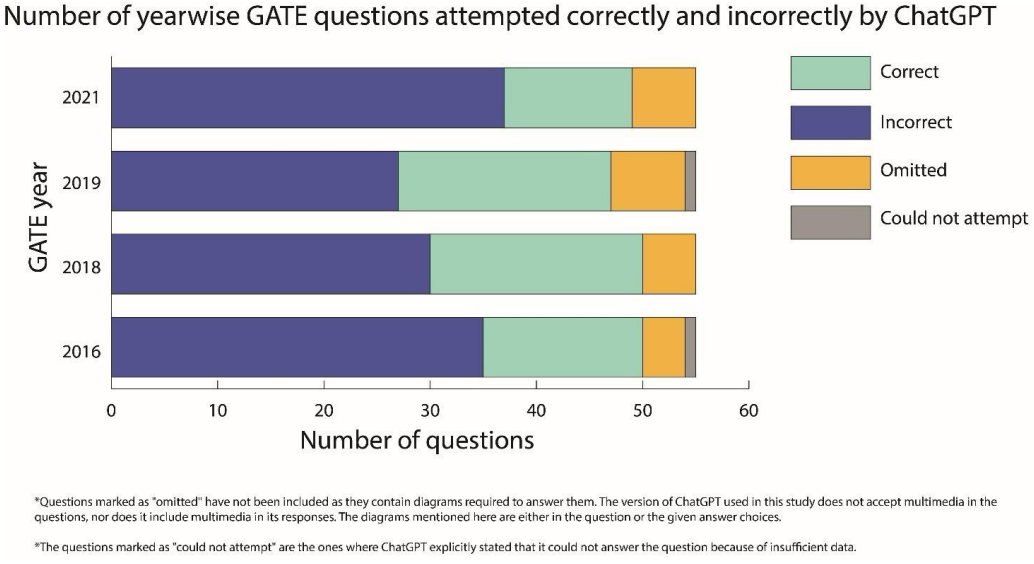
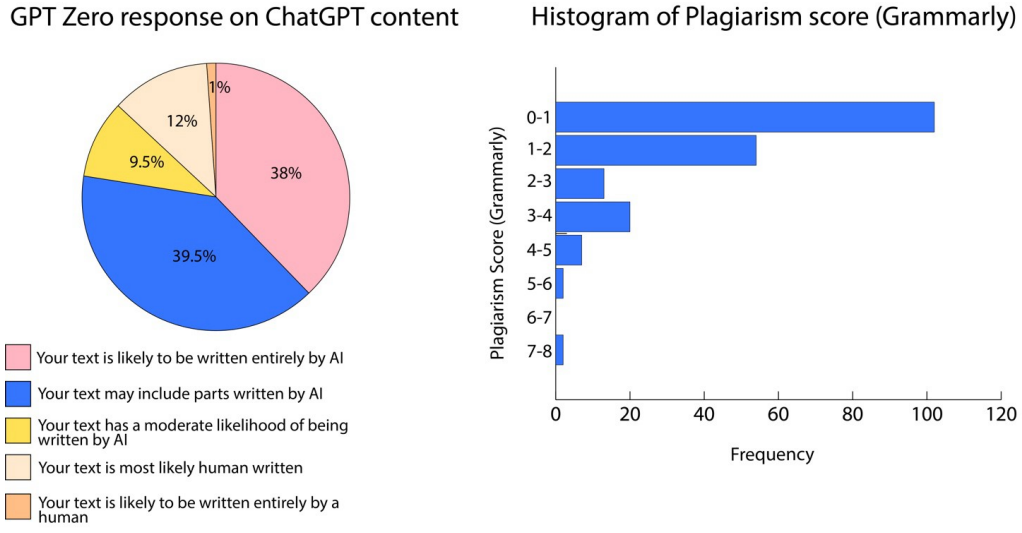


Figure 4 ChatGPT's performance in GATE Questions: Stacked bar graph illustrating correct and incorrect answers, highlighting its struggles in problem-Solving with a 20.4% accuracy rate.

ChatGPT's performance in solving GATE examination questions was evaluated, and it scored poorly with an average of 36.44% (Highest score – 41.6%, 2018; lowest score- 29.4%, 2021). The model

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

197 struggled particularly with questions that required problem-solving, such as numericals, achieving
198 only a 20.4% accuracy rate (Fig. 4; Supplementary file S3).
199
200 **3.2.2 Content Generation Performance Evaluation**
201
202 The content generation feature of ChatGPT emerged as the most utilized by geoscience students,
203 especially for generating content related to assignments, scripts, and during self-study sessions. To
204 evaluate the performance of this feature, two exercises were conducted, each focusing on different
205 aspects of content generation.



206
207 Figure 5 Assessing ChatGPT's Content Generation: A pie chart showcasing GPT Zero responses reveals the accuracy in detecting
208 AIgenerated content. Furthermore, a histogram of plagiarism scores for the same essays illustrates the low level of plagiarism in
209 ChatGPT's content.

210 ChatGPT was prompted to generate 200 essays covering various topics across different domains of
211 geosciences (Fig. 5; Supplementary file S4). The essays were subsequently analyzed for plagiarism
212 using Grammarly's built-in features. The results showed an average plagiarism rate of 1.46%,
213 indicating a low level of plagiarised content in the generated essays. Most essays had minimal or no
214 plagiarism, with 51% having less than 1% copied content and 94.5% having less than 4% copied
215 content. To further evaluate the authenticity of the generated content, GPTZero, a classification model

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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 AIgenerated 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 ChatGPT. The average improvement observed throughout the abstracts was 16.21%. An intriguing observation was that ChatGPT significantly improved the writing of poorly written texts (with low initial Grammarly Scores), thus following an exponential curve for improvement. All the abstracts that would not get improved later or showed minor improvements, originally had a score of more than 80, suggesting that the model is more effective in enhancing poorly written texts.

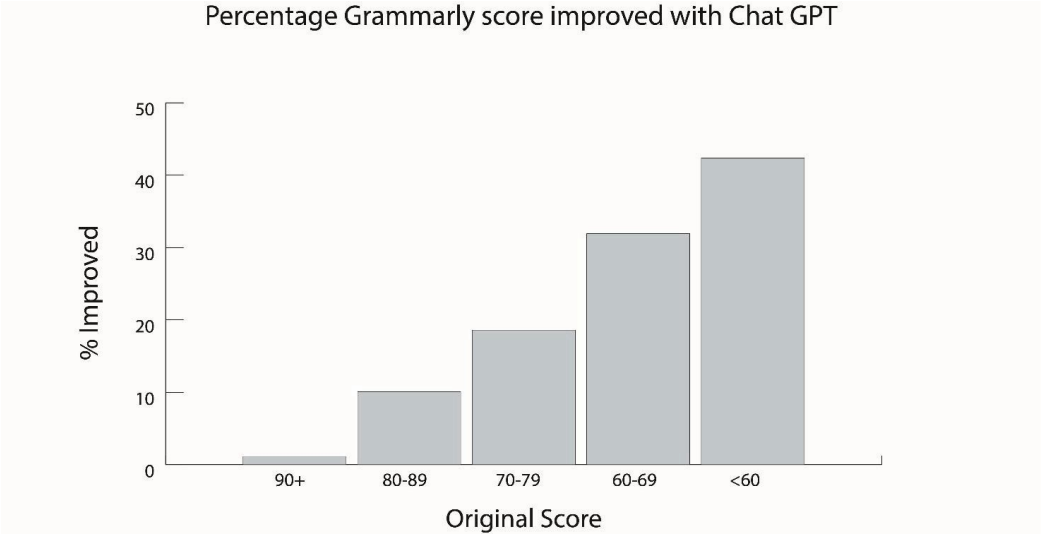


Figure 6 ChatGPT's impact on content quality enhancement. Analysis of Grammarly Scores before and after ChatGPT modifications reveals a 16.21% average improvement, with notable effectiveness in enhancing poorly written texts, as demonstrated by a significant improvement in previously low-scoring abstracts.

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

3.2.3 Translating Ability Assessment

To evaluate ChatGPT's translating service, we tested its ability to translate geological words and sentences from English to Hindi, which is commonly spoken in Mumbai and India. The translations were assessed using the metrics 'Accurate,' 'Moderate,' and 'Poor' to comment on the quality of the translations. Out of the geological words translated, 54% of them were accurately translated, meaning the Hindi translations were correct and aligned with their intended meanings (Supplementary file S5). However, a notable concern was that 26% of the translations were categorized as 'Poor,' indicating 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

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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 (Section 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

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

315 Permian-Triassic mass extinction and the impact as the cause of the fifth mass extinction in our
316 examples (Section 3.2.4). This bias could stem from the prevalence of certain topics in the training
317 data, possibly influenced by the availability of published literature and media coverage. , raising
318 concerns about the reliability of responses on certain topics. Future research could quantify data
319 sources to better understand and address bias in AI language models like ChatGPT.
320
321 Over-reliance on AI, including ChatGPT, may hinder the development of essential skills like critical
322 thinking, problem-solving, imagination, and research abilities in students. Worryingly, a considerable
323 percentage of students were unaware of the possibility of biased (46%), incorrect (47%), and outdated
324 (62%) responses from ChatGPT, highlighting the need for educational institutes to conduct awareness
325 sessions (Fig. 2). Promoting responsible usage and critical evaluation of AI language models will help
326 students harness the benefits while being mindful of the limitations.
327
328 **4.2 Pedagogical considerations**
329
330 The introduction of large language models like ChatGPT has ushered in a new era of technological
331 advancement in education. As technology continues to evolve rapidly, it inevitably impacts education
332 systems worldwide, prompting educators to explore the implications of incorporating AI technologies
333 into teaching and learning processes (Ausat et al., 2023). ChatGPT, as a powerful artificial intelligence
334 system capable of processing and generating sophisticated text, has the potential to revolutionize the
335 traditional classroom dynamic and raise critical questions about the role of teachers in the learning
336 process (Ausat et al., 2023; Fauzi et al., 2023).
337
338 Teachers play multifaceted roles beyond being instructors, serving as mentors and role models for
339 students (Zen et al., 2023). The introduction of ChatGPT and other AI technologies into the
340 geosciences educational landscape has the potential to complement and augment these roles in various

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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).

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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 (SalasPilco 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 humanmachine 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

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

consequences, as students using ChatGPT can obtain an unfair advantage over their peers who put in genuine efforts to produce original work (Bašić 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.

Beyond the issue of plagiarism, the implementation of AI in education also brings to light concerns about bias and inequalities. AI-assisted chatbots like ChatGPT can inadvertently perpetuate biases present in the training data, leading to the reinforcement of existing inequalities in education (Zhai, 2022). This raises ethical questions about ensuring fairness and equal opportunities for all students and highlights the need for developing AI systems that are free from inherent biases and discrimination. Another significant ethical consideration is the privacy and security of students' data. AI technologies collect and process vast amounts of data from users, including students, to improve their performance. However, there are valid concerns about how this data is utilized, stored, and protected. Safeguarding students' privacy and ensuring the secure handling of their data is of utmost importance to maintain trust in AI technologies in educational environments.

Furthermore, the potential for AI-generated content to contain errors or even fake information raises ethical questions regarding the dissemination of misinformation in scientific publications and academic work (Tlili et al., 2023; Liebrenz et al., 2023). While ChatGPT's responses are not exact copies of specific texts, their similarity to existing sources can lead to misleading content. This emphasizes the necessity of developing ethical guidelines for the use of AI in education to promote accuracy and credibility in academic work.

Despite these concerns, it is essential to acknowledge the positive aspects of integrating AI in education. ChatGPT and similar AI technologies have the potential to enhance instruction delivery

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

and learning practices, benefitting both teachers and students in various educational tasks, such as preparing teaching materials, creating quizzes, and offering personalized learning experiences (Kasneci et al., 2023). However, balancing the advantages with the ethical challenges is essential. The decision by New York City to ban ChatGPT in schools due to concerns about cheating in 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

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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 promoted, and educators must be proactive in upskilling to effectively leverage AI's benefits while mitigating its limitations.

In conclusion, ChatGPT holds promise in enhancing geosciences education, but its implementation should be done thoughtfully and responsibly. By understanding its capabilities and limitations, educators can leverage AI technologies to create more engaging, inclusive, and effective learning experiences for students while maintaining academic integrity and ethical standards.

Author Contribution

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

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

	Funding acquisition	Subham Patra

459

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

459
460

461 **Competing interests:**

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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.

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. The Head of Departments of the Earth Sciences dvision of participating institutes (IIT Bombay, K.J. Somaiya Mumbai, St. Xavier's College Mumbai) were duly informed, and necessary permissions for the study were obtained.

Acknowledgements:

Authors acknowledge the Department of Earth Sciences, IIT Bombay for letting us use assets necessary for carrying out this study. Authors thank Head of Departments of Earth Sciences division of participating institutes (IIT Bombay, K.J. Somaiya Mumbai, St. Xavier's College Mumbai) for letting us conduct the study. Authors also thank all the anonymous students who participated in the survey for this project.

References:

Ausat, A.M.A., Massang, B., Efendi, M., Nofirman, N. and Riady, Y., 2023. Can chat GPT replace the role of the teacher in the classroom: A fundamental analysis. Journal on Education, 5(4), pp.16100-16106. DOI: 10.31004/joe.v5i4.2745

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

Baidoo-Anu, D. and Owusu Ansah, L., 2023. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN 4337484.

Bašić, Ž., Banovac, A., Kružić, I. and Jerković, I., 2023. Better by You, better than Me? ChatGPT-3 as writing assistance in students' essays. DOI: 10.48550/arXiv.2302.04536

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

Choi, J.H., Hickman, K.E., Monahan, A. and Schwarcz, D., 2023. Chatgpt goes to law school. Available at SSRN. DOI: 10.2139/ssrn.4335905

Cooper, G., 2023. Examining science education in chatgpt: An exploratory study of generative artificial intelligence. Journal of Science Education and Technology, 32(3), pp.444-452. DOI: 10.1007/s10956-023-10039-y

Cotton, D.R., Cotton, P.A. and Shipway, J.R., 2023. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, pp.1-12. DOI: 10.1080/14703297.2023.2190148

Dong, Y. and Shi, L., 2021. Using Grammarly to support students' source-based writing practices. Assessing Writing, 50, p.100564. DOI: 10.1016/j.asw.2021.100564

Elbanna, S. and Armstrong, L., 2023. Exploring the integration of ChatGPT in education: adapting for the future. Management & Sustainability: An Arab Review. DOI: 10.1108/MSAR-03-2023-0016

Farrokhnia, M., Banihashem, S.K., Noroozi, O. and Wals, A., 2023. A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, pp.1-15. DOI: 10.1080/14703297.2023.2195846 CrossMark

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A.M.A. and Hatta, H.R., 2023. Analysing the role of ChatGPT in improving student productivity in higher education. Journal on Education, 5(4), pp.14886-14891. DOI: 10.31004/joe.v5i4.2563

Gilson, A., Safranek, C.W., Huang, T., Socrates, V., Chi, L., Taylor, R.A. and Chartash, D., 2023. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. JMIR Medical Education, 9(1), p.e45312. DOI: 10.2196/45312

Haleem, A., Javaid, M., Qadri, M.A. and Suman, R., 2022. Understanding the role of digital technologies in education: A review. Sustainable Operations and Computers, 3, pp.275-285. DOI: 10.1016/j.susoc.2022.05.004

Hargreaves, S., 2023. 'Words Are Flowing Out Like Endless Rain Into a Paper Cup': ChatGPT & Law School Assessments. The Chinese University of Hong Kong Faculty of Law Research Paper, (2023-03). DOI: 10.2139/ssrn.4359407

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

Hwang, G.J. and Chang, C.Y., 2021. A review of opportunities and challenges of chatbots in education. Interactive Learning Environments, pp.1-14. DOI: 10.1080/10494820.2021.1952615

Jiao, W., Wang, W., Huang, J.T., Wang, X. and Tu, Z.P., 2023. Is ChatGPT a good translator? Yes with GPT-4 as the engine. arXiv preprint arXiv:2301.08745. DOI: 10.48550/arXiv.2301.08745

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E. and Krusche, S., 2023. ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, p.102274. DOI: 10.1016/j.lindif.2023.102274

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

526 Keller, G., Mateo, P., Monkenbusch, J., Thibault, N., Punekar, J., Spangenberg, J.E., Abramovich, S.,
527 Ashckenazi-Polivoda, S., Schoene, B., Eddy, M.P. and Samperton, K.M., 2020. Mercury linked to
528 Deccan Traps volcanism, climate change and the end-Cretaceous mass extinction. Global and
529 Planetary Change, 194, p.103312. DOI: 10.1016/j.gloplacha.2020.103312

530 Khalil, M. and Er, E., 2023. Will ChatGPT get you caught? Rethinking of plagiarism detection. arXiv
531 preprint arXiv:2302.04335. DOI: 10.48550/arXiv.2302.04335

532 Li, L., Ma, Z., Fan, L., Lee, S., Yu, H. and Hemphill, L., 2023. ChatGPT in education: A discourse
533 analysis of worries and concerns on social media. arXiv preprint arXiv:2305.02201. DOI:
534 10.48550/arXiv.2305.02201

535 Liebrenz, M., Schleifer, R., Buadze, A., Bhugra, D. and Smith, A., 2023. Generating scholarly content
536 with ChatGPT: ethical challenges for medical publishing. The Lancet Digital Health, 5(3),
537 pp.e105e106. DOI: 10.1016/S2589-7500(23)00019-5

538 Lo, C.K., 2023. What is the impact of ChatGPT on education? A rapid review of the literature.
539 Education Sciences, 13(4), p.410. DOI: 10.3390/educsci13040410

540 Salas-Pilco, S.Z., Xiao, K. and Hu, X., 2022. Artificial intelligence and learning analytics in teacher
541 education: A systematic review. Education Sciences, 12(8), p.569. DOI: 10.3390/educsci12080569
542 Sallam, M., 2023. The utility of ChatGPT as an example of large language models in healthcare
543 education, research and practice: Systematic review on the future perspectives and potential
544 limitations. medRxiv, pp.2023-02. DOI: 10.1101/2023.02.19.23286155

545 Schulte, P., Alegret, L., Arenillas, I., Arz, J.A., Barton, P.J., Bown, P.R., Bralower, T.J., Christeson,
546 G.L., Claeys, P., Cockell, C.S. and Collins, G.S., 2010. The Chicxulub asteroid impact and mass
547 extinction at the Cretaceous-Paleogene boundary. Science, 327(5970), pp.1214-1218. DOI:
548 10.1126/science.1177265

https://doi.org/10.5194/gc-2023-7
Preprint. Discussion started: 3 January 2024

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

552 Tlili, A., Shehata, B., Adarkwah, M.A., Bozkurt, A., Hickey, D.T., Huang, R. and Agyemang, B.,
553 2023. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education.
554 Smart Learning Environments, 10(1), p.15. DOI: 10.1186/s40561-023-00237-x

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

558 Yang, M., 2023. New York City schools ban AI chatbot that writes essays and answers prompts. The
559 Guardian, 6.
560 Zawacki-Richter, O., Marín, V.I., Bond, M. and Gouverneur, F., 2019. Systematic review of research
561 on artificial intelligence applications in higher education–where are the educators?. International
562 Journal of Educational Technology in Higher Education, 16(1), pp.1-27. DOI: 10.1186/s41239-
563 0190171-0

564 Zen, A., Kusumastuti, R., Metris, D., Gadzali, S.S. and Ausat, A.M.A., 2023. Implications of
565 Entrepreneurship Education as a Field of Study for Advancing Research and Practice. Journal on
566 Education, 5(4), pp.11441-11453. DOI: 10.31004/joe.v5i4.2091

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