

## Is ChatGPT reliable in education?

Amal Abdullah Alibrahim 

Department of Curriculum and Instruction, College of Education, King Saud University, Riyadh, Saudi Arabia  
amabdull@ksu.edu.sa

After ChatGPT was released late in 2022, many arguments about its accuracy and use in education arose. In this article, I seek to provide evidence of the accuracy and validity of ChatGPT's responses to users' queries in education by applying a systematic review methodology to analyse publications in specific databases following PRISMA guidelines which provide a high level of evidence. Of 274 publications initially identified, 18 were included based on eligibility criteria. My findings show some limitations of ChatGPT, for example, a lack of deep understanding, limited ability to calculate problems, and difficulty with complex problems. Despite these limitations it was clear that ChatGPT was able to pass many exams and succeed in many assessment problems in a variety of education disciplines. Finally, based on the findings, I suggest an ABCD framework to successfully apply ChatGPT in education.

**Keywords:** AI; authenticity of AI in education; ChatGPT; LLM

### Introduction

Chat generative pre-trained transformer (ChatGPT), developed by OpenAI, has gained considerable attention as it can generate human-like text and has, since the release of ChatGPT-3.5 in late November 2022, been introduced in many fields, e.g. medicine, business, and education. The first generation of the GPT model, GPT-1, was introduced in June 2018 as a language understanding model. The model learns and improves from user interaction. ChatGPT-2 in which more parameters and data sets were used was released in February 2019. In 2020, GPT-3 was introduced with increasing parameters – all based on a GPT one-way language model training method. OpenAI update GPT-3.5 at the end of November 2022, based on the dialogue mode, simulating human dialogue and thinking. With accuracy increased by 40%, ChatGPT-4 was released in March 2023 as a large language model (LLM) (OpenAI, 2023; Wang, 2024). A summary of the development of ChatGPT is presented in Table 1. The software's capacity to mimic humans had been significantly improved. ChatGPT-3.5 was trained on a massive quantity of data from the internet up to October 2021 and carefully curated to provide a diverse and representative sample of human language. These data include sources such as books, articles, websites, and other publicly available documents (ChatGPT.com, n.d.). It has the ability to self-learn (Farrokhnia, Banihashem, Noroozi & Wals, 2024). It is remarkable how fast its growth has impacted education; in less than a week following its initial public launch, ChatGPT attracted over a million users, with more than 100 million currently active users (Baidoo-Anu & Owusu Ansah, 2023), while other popular platforms such as Facebook, Netflix, Instagram, Twitter (X) took much longer to reach 1,000,000 users, (300, 1,200, 75, and 720 days) respectively (Biswas, 2023).

**Table 1** Summary of GPT development

Generation	Date of release	Main features
GPT1	June 2018	Language understanding model
GPT2	February 2019	More parameters and data sets
GPT3	2020	Increased parameters
GPT3.5	November 2022	Dialogue mode
GPT 4	March 2023	Increased accuracy rate

Many scholars have studied ChatGPT's potential use and impact in education. Its ability to understand questions in natural language and respond in a coherent and contextually relevant manner implies its potential for use in learning languages, communication, and education (Baskara & Mukarto, 2023). A large number of students are familiar with and have a positive attitude towards ChatGPT in education, and they acknowledge its potential use in their studies; nevertheless, they do not consistently use it for their studies, which emphasises the need to instruct students of appropriate use to support their learning (Lozano & Fontao, 2023; Singh, H, Tayarani-Najaran & Yaqoob, 2023). Furthermore, teachers believe that using ChatGPT in higher education will support their students and improve learning and teaching, and they are willing to use it in their teaching (Chan, 2023).

The introduction of ChatGPT to the educational field has generated conflicting reactions among educators, as it may change current educational practices dramatically (Baidoo-Anu & Owusu Ansah, 2023). Integrating ChatGPT into education raises many concerns about plagiarism, cheating in assignments, and impacts on critical thinking, especially considering that ChatGPT's writing cannot be detected by plagiarism detection software (e.g., Turnitin).

ChatGPT has gained significant interest from scholars as it has a significant impact on education. Many

studies present the impact of using ChatGPT in the education field and describe the challenges, opportunities, and concerns (Farrokhnia et al., 2024; Michel-Villarreal, Vilalta-Perdomo, Salinas-Navarro, Thierry-Aguilera & Gerardou, 2023; Zeb, Ullah & Karim, 2024). It is claimed that using ChatGPT has many potential advantages, including improving student learning (Netto, 2023) and enhancing collaboration, student engagement, and accessibility (Cotton, Cotton & Shipway, 2024; Zeb et al., 2024). Nevertheless, new learning settings and teaching strategies should be applied, and students guided to use it acceptably in their learning (Singh, M 2023; Zeb et al., 2024).

ChatGPT can personalise learning and create authentic course material (Baskara & Mukarto, 2023). In addition, ChatGPT can be an effective teacher assistant as it can help to analyse course outcomes, provide personal feedback and advice, and create assessment questions (Bonner, Lege & Frazier, 2023), which can lead to innovative and authentic assessment (Crawford, Cowling & Allen, 2023). ChatGPT has also demonstrated its potential benefit in the teaching/learning of English (Shaikh, Yayilgan, Klimova & Pikhart, 2023). Students' critical and creative thinking skills developed while giving ChatGPT instructions in analysis tasks (Shue, Liu, Li, Feng, Li & Hu, 2023). Furthermore, when allowed creativity to interact with ChatGPT, students could acquire critical and creative skills (Ellis & Slade, 2023; Javier & Moorhouse, 2023; Zeb et al., 2024). In addition, the integration of ChatGPT in educational settings can enhance students' decision-making skills. How the use of ChatGPT has improved investment decisions in the Pakistan stock market by aiding in data analysis and risk management (Ullah, Ismail, Khan & Zeb, 2024), is an example. It acts as a student support service and gives personalised feedback (Chan, 2023). Also, ChatGPT can be treated as a tutoring system by helping students to develop critical thinking and debating skills (Farrokhnia et al., 2024). Consequently, educators are open to using ChatGPT in education (Singh, M 2023), while they acknowledge its advantages and limitations (Chan, 2023).

On the other hand, many challenges and limitations of ChatGPT have been reported. Some of the main challenges are a lack of deep understanding and impact on students' higher-order thinking, biased results, and the lack of accuracy and reliability (Farrokhnia et al., 2024; Karabacak, Ozkara, Margetis, Wintermark & Bisdas, 2023; Michel-Villarreal et al., 2023; Su & Yang, 2023; Zhu, Sun, Luo, Li & Wang, 2023). Also, ChatGPT could give false information with confidence (Khosravi, Al Sudani & Oladnabi, 2023).

Furthermore, using ChatGPT in education has created many concerns about cheating and plagiarism, as it is difficult to determine whether

written text is human or machine-generated, which may detract from independent and higher-order thinking (Chan, 2023; Cotton et al., 2024; Ellis & Slade, 2023; Farrokhnia et al., 2024; Yang & Stivers, 2024; Zeb et al., 2024). This is especially challenging as plagiarism detection software cannot detect ChatGPT writing (Chaudhry, Sarwary, El Refae & Chabchoub, 2023). Many suggestions have been made to overcome these challenges, e.g., changing the way in which assessment is performed (Kelly, Sullivan & Strampel 2023; Singh, M 2023; Zeb et al., 2024). Many studies emphasise developing new policies and establishing guidelines to use ChatGPT in education (Chan, 2023; Karabacak et al., 2023; Michel-Villarreal et al., 2023; Zeb et al., 2024).

OpenAI, in their technical report, claim that ChatGPT can accept image and text prompts and answer in text. Also, it can perform in many academic exams, scoring within the top 10% (OpenAI, 2023). However, studies report that the accuracy of ChatGPT cannot be guaranteed. Its ability to respond correctly and accurately and show conceptual understanding is called into question in different disciplines.

In addition, with the numerous uses of ChatGPT in education, it is inevitable to evaluate the accuracy and validity of its responses. The aim with the study reported on here was to assess the accuracy and validity of ChatGPT answers in education. I assessed ChatGPT's responses to user prompts through empirical research, building on previous studies that highlight the need for further investigation into the potential impact of introducing ChatGPT into an educational setting (Michel-Villarreal et al., 2023).

My study makes several contributions to the existing knowledge. Firstly, the findings of the study will benefit institutions and educators by suggesting actions needed to successfully use ChatGPT in education. Secondly, provides teachers and students with a benchmark and clear vision for the accuracy and validity of ChatGPT responses based on empirical studies. Thirdly, the findings of this study could inform developers regarding the improvement needed for next generations of ChatGPT and reveal weak points for improvement.

## Method

Research on ChatGPT has increased rapidly; many studies have validated the accuracy of ChatGPT from different perspectives (e.g., thin ethnographic research involving ChatGPT [Michel-Villarreal et al., 2023; Stojanov, 2023], writing a research paper [Cotton et al., 2024], and SWOT [strengths, weaknesses, opportunities, and threats] analyses [Farrokhnia et al., 2024]). The main aim with this study was to collect empirical evidence from studies in which the accuracy and validity of ChatGPT responses were evaluated by applying a

systematic review to integrate the findings from empirical studies. The findings of this study are reliable, as it addresses research questions that remain unexplored by individual studies. This is achieved through the methodological rigor of a systematic review, which synthesises evidence from multiple studies. Systematic review methods are designed to comprehensively identify all empirical studies that meet predefined inclusion criteria, ensuring a thorough and unbiased examination of the research questions (Snyder, 2019). The systematic review approach has a high level of evidence (Tawfik, Dila, Mohamed, Tam, Kien, Ahmed & Huy, 2019). Snyder (2019) claims that the systematic review is suitable for collecting

evidence. In accordance with the guidelines for the Preferred reporting items for systematic reviews and meta-analyses (PRISMA), a systematic review method was carried out in this study (PRISMA, 2023).

**Information Sources and Search Strategy**

After the research question had been identified, a replicable search strategy was developed. Inclusion and exclusion criteria were set to select eligibility for the study, after which an electronic search was carried out using selected search terms on many databases, including Web of Science, EBSCOhost, and Google Scholar. The search strategy is illustrated in Table 2.

**Table 2** Search strategy

Databases (providers)	Number (No.) of studies	Search terms
Google Scholar	101	ChatGPT + Education
Web of science	103	ChatGPT + Education
Results from Social Sciences Citation Index (SSCI), Conference Proceedings Citation Index – Science (CPCI-S), Conference Proceedings Citation Index – Social Science & Humanities (CPCI-SSH), Emerging Sources Citation Index (ESCI).		
EBSCO	70	ChatGPT + Education

**Eligibility Criteria**

Based on the aim to evaluate ChatGPT responses in education, eligibility criteria were defined, which

are shown in Table 3.

**Table 3** Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Peer review	Unrelated subject
Empirical study	Full text was not available (abstract only)
English	Duplication

**Study Selection**

After searching the selected databases, the collected studies were checked, and duplicate studies were removed. Next, a review and a preliminary screening were performed, which involved reading the title and abstract and selecting studies based on eligibility criteria. If there were doubts about a study’s inclusion, it was considered for further screening. Thereafter, the full studies were screened to verify whether they met the inclusion criteria. Studies that focused on and were related to the research aim and satisfied the inclusion criteria were included in the study.

**Coding, Data Extraction, and Analysis**

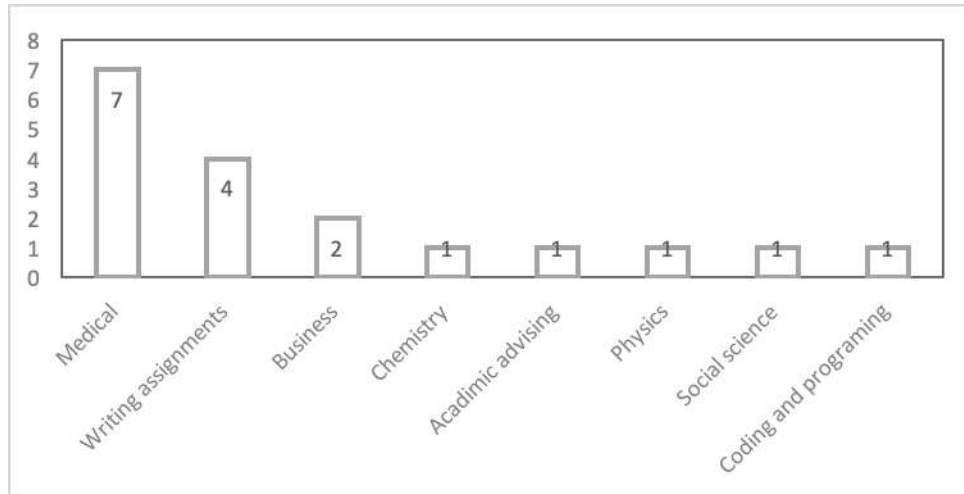
After having selected the studies, 18 studies were reviewed. To analyse the data and extract codes, the eligible studies were exported to a Microsoft

Excel spreadsheet. A thematic analysis was used to extract the codes and themes.

**Results**

As ChatGPT is used extensively in education sectors, it is essential to evaluate the accuracy and credibility of its responses. As this was the main aim of this study, I collected evidence from research in many disciplines in the education field.

An analysis of the scholarly sources selected for the study shows that ChatGPT has been applied in many fields, mainly the medical field, writing assignments, business, finance problems, chemistry, academic advising, physics, social science, and coding and programming. The number of published studies in each field is shown in Figure 1 and the papers from each field included in this study are displayed in Table 4.



**Figure 1** Number of published studies based on fields

**Table 4** Papers included

Field	Papers
Medical	Borchert, Hickman, Pepys & Sadler, 2023 Friederichs, Friederichs & März, 2023 Huang, Goma, Semrau, Haderlein, Lettmaier, Weissmann, Grigo, Tkhatay, Frey, Gaipi, Distel, Maier, Fietkau, Bert & Putz, 2023 Huh, 2023 Khosravi et al., 2023 Lahat, Shachar, Avidan, Glicksberg & Klang, 2023
Writing assignments	Morjaria, Burns, Bracken, Ngo, Lee, Levinson, Smith, Thompson & Sibbald, 2023 Farazouli, Cerratto-Pargman, Bolander-Laksov & McGrath, 2024 Tirado-Olivares, Navío-Inglés, O'Connor-Jiménez & Cózar-Gutiérrez, 2023 Tülübaş, Demirkol, Ozdemir, Polat, Karakose & Yirci, 2023 Waltzer, Cox & Heyman, 2023
Business	Chaudhry et al., 2023
Finance problems	Yang & Stivers, 2024
Chemistry	Daher, Diab & Rayan, 2023
Academic advising	Akiba & Fraboni, 2023
Physics	Kortemeyer, 2023
Social science	Netto, 2023
Coding and programming	Ouh, Gan, Shim & Wlodkowski, 2023

From the analysis of the articles that were included, two main categories were identified: the first determines the quality of ChatGPT's writing and the quality of the generated text, and the second assesses its responses as a learner and its accuracy and validity in the scientific field.

#### ChatGPT Writing

From an evaluation and analysis of the different articles included in the study, ChatGPT demonstrated that it can generate good, coherent text of which the quality of writing is good, despite reports in many articles that ChatGPT fails in critical thinking. Surprisingly, one article reports that it is authentic and generates creative writing (Farazouli et al., 2024). One study highlights that ChatGPT can write essays at a higher quality than most high school students (Waltzer et al., 2023). Making similar observations and arguments, ChatGPT is efficient in writing historical essays (Tirado-Olivares et al., 2023). Other characteristics

of ChatGPT responses and its quality of writing are highlighted in other studies, where ChatGPT provided answers for assignments written with a good structure that were very precise, covered all exam points, were concise, and provided arguments. Some answers were creative and innovative. However, the answers sometimes lacked argumentation and related content (Farazouli et al., 2024).

As a language generator, ChatGPT showed a clear, comprehensive, and deep response regarding open-ended career-related questions; while answering students as a college teacher, ChatGPT's answers were high-quality, confidential, and supportive. Some ChatGPT answers were of higher quality than those any adviser could provide. It demonstrated that it could act as a human adviser for elementary school teaching (Akiba & Fraboni, 2023). When ChatGPT answered a case study assessment, it was clear, accurate, correct, and realistic; on the other hand, the answers lacked

depth of understanding and did not provide details related to the case study (Netto, 2023). It is notable that ChatGPT responses change based on query keywords (Chaudhry et al., 2023). However, it was difficult to distinguish between essays written by ChatGPT and high school students, as high school teachers and students report that while the accuracy was 70% and 62% for teachers and students, respectively, in identifying who had written the essays, the percentage of guessing was high. One explanation was that they guessed that high-quality writing was achieved by ChatGPT, as reported (Waltzer et al., 2023).

#### Chat GPT's Performance as a Learner

Furthermore, from the analysis of the selected articles, ChatGPT can pass many exams at undergraduate level. Other studies have highlighted the academic performance of ChatGPT, and it is indicated that ChatGPT can pass a variety of instruments used for evaluating undergraduate students' learning objectives in many business courses (Chaudhry et al., 2023). ChatGPT is capable of writing assignments based on historical thinking at a higher standard than university students (Tirado-Olivares et al., 2023).

Moreover, from reviewing included studies, there is evidence that ChatGPT could perform scientific research, producing clear, accurate, concise, and unbiased information, although responses lack depth and are relatively repetitive (Tülübaş et al., 2023). ChatGPT as an undergraduate student is discussed in other studies demonstrating how ChatGPT can perform like a Grade A student. ChatGPT took assignments from five courses in business with different forms of assessment and different complexity levels. Its response to the case study was excellent: coherent, critical, and good communication free from language errors was demonstrated. Its answer for self-reflection work was excellent; critical, well-thought, sufficient and well-communicated responses were provided. There was, however, a lack of recommendations and in-depth analysis in project assignments. In problem solving, it answered well as the analysis was comprehensive, covering all necessary procedures with a sufficient justification for the recommendations made (Chaudhry et al., 2023). Testing the ability of ChatGPT to solve calculation-based problems within finance undergraduate courses showed that ChatGPT faced difficulty in answering advanced problems, where it answered 85% questions correctly from the basic problems of finance courses, 20% from the medium difficulty level, and 12% from the hard difficulty level (Yang & Stivers, 2024).

Also, as cited from the analysis, ChatGPT has revealed its promising ability to pass many tests in humanities, social sciences, and undergraduate law

courses with grades over 66% (Farazouli et al., 2024). The findings from the articles analysed also demonstrate that ChatGPT can pass the Radiation Oncology In-Training (TXIT) exam with a score of 78.77% but lacks thorough details of clinical trials (Huang et al., 2023). ChatGPT succeeded in the Situational Judgement Test (SJT – United Kingdom), with a score of 76% (Borchert et al., 2023). For the Red Journal Gray Zone cases, it demonstrated accurate and comprehensive treatment for each case. In some cases, it provided creative treatment suggestions (Huang et al., 2023). On the other hand, as claimed from other articles analysed, ChatGPT shows a lack of understanding of the gastroenterology field; its answers to prompted questions were lacking in either accuracy or completion or showed a lack of understanding in the subject (Lahat et al., 2023). ChatGPT obtained a score of 65.5% in the test, performing better than all medical students in years 1 to 3 (Friederichs et al., 2023). On the other hand, ChatGPT's performance in a parasitology examination was worse than that of Korean medical students, with scores of 60.8% while students' average score was 77%. One of the explanations is that some exam materials (epidemiological data) are specific to Korea (Huh, 2023). In another case, ChatGPT could pass a short-answer test in a pre-clerkship medical educational programme (mean 3.29); however, its performance was low compared to the students' answers with a mean of 3.67 (Morjaria et al., 2023). Furthermore, ChatGPT can pass a Doctor of Philosophy (Ph.D.) entrance exam in medical genetics, scoring 52/80, exhibiting good performance with 70% correct answers, but it struggled with analytical and critical thinking (Khosravi et al., 2023).

In the science discipline, ChatGPT could solve 13 out of 30 chemistry problems (43.33%). ChatGPT has difficulty in chemistry problems, especially representations and depth and synthesis, evaluation, and analysis problems. On the other hand, all memorised problems were solved correctly (Daher et al., 2023). ChatGPT can pass a physics course (1.5 out of 4.0), acting like a beginner learning physics with some errors (Kortemeyer, 2023). Also, ChatGPT can write a program in Java, with high readability and well-structured code, and it can suggest alternative solutions to increase memory efficiency (Ouh et al., 2023).

Furthermore, as highlighted from the different sources reviewed and analysed, some of ChatGPT's responses were incorrect, and it gives false information as if it were true with complete confidence (Yang & Stivers, 2024). Also, there is evidence in some cases that some answers were hallucinations (Huang et al., 2023). Other studies also discussed that some of its responses were lacking in depth (Tülübaş et al., 2023).

ChatGPT answers sometimes lack argumentation and related content (Farazouli et al., 2024). Some studies analysed highlighted that some of ChatGPT's responses were relatively repetitive (Tülübaş et al., 2023). Lack of judgment and reasoning (Borchert et al., 2023) were also noted, along with some difficulty in calculating formulas within square roots (Kortemeyer, 2023). ChatGPT cannot understand prompts with nuances and provides incorrect answers with confidence (Khosravi et al., 2023). As ChatGPT is a natural language processing model, it fails to answer coding questions with non-textual descriptions (Ouh et al., 2023). In a study where ChatGPT was asked the same question twice, it did not provide the same answer, nor even the same result in terms of correctness (Kortemeyer, 2023).

### Discussion

The development of ChatGPT has influenced many fields to apply and use it. Many studies discuss the potential affordances of ChatGPT and its significant impact on a variety of fields. The education field is one where ChatGPT has been applied and has had a significant impact. With the rapid use and growth of user numbers among students and instructors alike, the validity and accuracy of its responses need to be evaluated and assessed. The aim with this study was to evaluate the accuracy and validity of ChatGPT responses by applying a systematic review.

It is clear from the studies mentioned above that ChatGPT potentially has an important impact on education; however, ChatGPT should be used with thorough consideration of its responses and their validation. ChatGPT responses perform well in some exams and can pass many undergraduate exams. Nevertheless, its performance in the scientific field is limited, as its main purpose is to produce and generate human-like text.

It is important to highlight the fact that ChatGPT is a text-generation tool, and different studies have discussed and demonstrated the quality of ChatGPT's writing responses, which were found to be coherent, correct, well-structured, precise, concise, and free from language errors.

On the other hand, ChatGPT responses were found to be lacking in argumentation and unrelated to the content, as well as lacking depth of understanding and higher-order thinking.

The studies analysed agree that ChatGPT can perform well as a beginner student in undergraduate courses and makes the same errors as the students, and it exceeds in exams that are based on writing essays. However, ChatGPT struggles to answer scientific, complex problems.

In addition, it is demonstrated that ChatGPT has limitations with calculation prompts, deep thinking, deep understanding of complex subjects, showing some issues in its responses. Despite this,

ChatGPT responses in some cases were unusual or novel, which counts as a creative answer (Farazouli et al., 2024). The same is highlighted in a study by Huang et al. (2023) where ChatGPT showed novel treatment approaches in some cases. Nevertheless, these cases were limited, and a repeated answer from the prompt is not guaranteed, as is evident in Kortemeyer's study (2023) where the answer from ChatGPT differed every time when the same prompt text was entered. Other issues regarding ChatGPT were that it sometimes answers in hallucinations and cannot recognise human nuances in an image prompt, while the new update of ChatGPT reports that it can accept image and text prompts (OpenAI, 2023). In addition, one of the limitations that ChatGPT shows in education is calculation problems, namely calculating equations correctly. As a language model, it may perform calculations using sophisticated pattern-matching rather than processing mathematics equations (Kortemeyer, 2023).

One reason for the inaccuracy and unreliability of some of ChatGPT's responses is that its most recent training was completed in late 2021, and its responses are based on data sourced from the internet. Also, its main purpose is to generate human-like text responses. Nevertheless, its potential impact in education should be acknowledged, and its future is a cause for optimism.

### Implications of the Findings

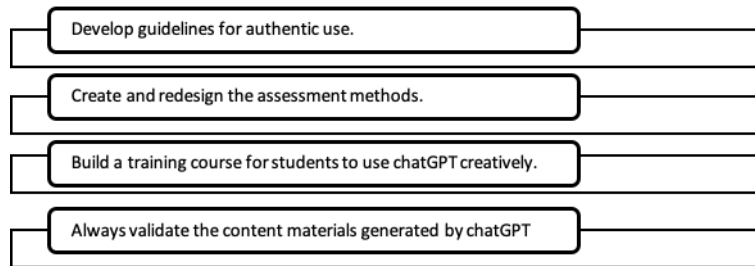
Based on the discussion above, some of the implications of my research which could benefit educational institutions and their members to adopt this new technology in their practice successfully is discussed below. Educational institutions should keep adapting and aligning with the dynamic digital society to remain connected and relevant. The practical implications of this research are as follows.

- Although ChatGPT responses in some cases were coherent and sensible, its accuracy in its responses is limited for answering memorised problems and writing academic essays, and it struggles with complex scientific problems. Validation of the content of ChatGPT responses should be required to guarantee data accuracy and reliability. This echoes many researchers' recommendations (Huang et al., 2023; Tülübaş et al., 2023).
- As a text-based generator model, the keyword prompt for ChatGPT should be text based and carefully selected and structured to guarantee the accuracy and relevance of the response materials. Despite this, it is claimed that ChatGPT can understand images (OpenAI, 2023).
- Educators should redesign the assessment scheme to prevent cheating or use of ChatGPT in dishonest ways. For example, teachers can use topics and assignments that measure higher-order skills and provide some guidelines for how to interact with ChatGPT to develop an education objective, as

- many educators advocate (e.g. Singh, M 2023).
- An authentic assessment can be designed around ChatGPT that helps students develop and improve critical and higher-order thinking skills.
- Education institutions should develop guidelines and policies for the acceptance and use of ChatGPT.
- Students should be guided on how to use ChatGPT

constructively and trained to use ChatGPT in their learning.

Clear consideration in adopting ChatGPT in education is important. To adopt ChatGPT successfully, I propose the “ABCD framework” shown in Figure 2, which includes steps for education practitioners to follow.



**Figure 2** ABCD framework to adopt ChatGPT in education

### Conclusion

As a result of the extensive use of ChatGPT following its release in late 2022, and with its potential use in education, it is essential to assess the validity and accuracy of ChatGPT responses in education sectors, which was the main purpose with this study.

My findings lead to the conclusion that ChatGPT demonstrates its potential use in education in many disciplines, although ChatGPT responses were not fully correct or relevant, especially within the scientific field. While ChatGPT shows that it can pass many exams, its answers lack deep understanding of the subject matter, and higher-order thinking skills. In addition, the findings included analyses of studies highlighting the need to enhance ChatGPT responses to complex scientific problems and calculation problems. It is predicted that the new generation of ChatGPT will learn and update its data, which may result in improvement over time as ChatGPT acquires more data. In addition, cooperation between educational institutions and developers to adapt a special ChatGPT for education domains is suggested. It is important to acknowledge the fast improvement and development of ChatGPT.

Note that during the editing of this research, OpenAI released a new update to ChatGPT (ChatGPT-4), with many different capabilities and improvements (OpenAI, n.d.) like new voice and image capabilities (released on 21 November 2023 for all users), browsing the internet for new updated information with sources (released on September 27, 2023), and custom GPTs. Custom GPTs are targeted at developers who would like to tailor and explore the potential use of a

personalised assistant. This powerful tool could be used to train a custom GPT on a specific topic using textbooks, for example, and for use as a teacher assistant where it would be able to prepare lessons, and answer student questions. Further research that addresses customised ChatGPT for a course and for learning and developing students' skills is needed for all disciplines.

### Limitations and Further Studies

Some limitations of this study should be highlighted here. The systematic review focused on specific databases, and the search process was applied up to 26 October 2023 and was limited to studies that I could access. In addition, only studies that met the eligibility criteria (e.g., empirical and peer reviewed studies) were analysed.

Finally, future research should consider the potential effects of using AI (especially ChatGPT) on students' learning. Moreover, it is essential for future research to design and examine innovative frameworks for students' evaluation.

### Notes

- Published under a Creative Commons Attribution Licence.
- DATES: Received: 30 March 2024; Revised: 26 October 2024; Accepted: 19 November 2024; Published: 30 November 2024.

### References

- Akiba D & Fraboni MC 2023. AI-supported academic advising: Exploring ChatGPT's current state and future potential toward student empowerment. *Education Sciences*, 13(9):885. <https://doi.org/10.3390/educsci13090885>
- Baidoo-Anu D & Owusu Ansah L 2023. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in

- promoting teaching and learning. *Journal of AI*, 7(1):52–62. <https://doi.org/10.61969/jai.1337500>
- Baskara R & Mukarto M 2023. Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 7(2):343–358. <https://doi.org/10.21093/ijeltal.v7i2.1387>
- Biswas S 2023. *Role of Chat GPT in education*. Available at <https://ssrn.com/abstract=4369981>. Accessed 12 November 2024.
- Bonner E, Lege R & Frazier E 2023. Large Language Model-based Artificial Intelligence in the language classroom: Practical ideas for teaching. *Teaching English with Technology*, 23(1):23–41. <https://doi.org/10.56297/BKAM1691/WIEO1749>
- Borchert RJ, Hickman CR, Pepys J & Sadler TJ 2023. Performance of ChatGPT on the Situational Judgement Test—A professional dilemmas-based examination for doctors in the United Kingdom. *JMIR Medical Education*, 9:e48978. <https://doi.org/10.2196/48978>
- Chan CKY 2023. A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20:38. <https://doi.org/10.1186/s41239-023-00408-3>
- ChatGPT.com n.d. Available at <https://chat.openai.com>. Accessed 16 November 2023.
- Chaudhry IS, Sarwary SAM, El Refae GA & Chabchoub H 2023. Time to revisit existing student's performance evaluation approach in higher education sector in a new era of ChatGPT — A case study. *Cogent Education*, 10(1):2210461. <https://doi.org/10.1080/2331186X.2023.2210461>
- Cotton DRE, Cotton PA & Shipway JR 2024. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2):228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Crawford J, Cowling M & Allen KA 2023. Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching and Learning Practice*, 20(3):1–19. <https://doi.org/10.53761/1.20.3.02>
- Daher W, Diab H & Rayan A 2023. Artificial intelligence generative tools and conceptual knowledge in problem solving in chemistry. *Information*, 14(7):409. <https://doi.org/10.3390/info14070409>
- Ellis AR & Slade E 2023. A new era of learning: Considerations for ChatGPT as a tool to enhance statistics and data science education. *Journal of Statistics and Data Science Education*, 31(2):128–133. <https://doi.org/10.1080/26939169.2023.2223609>
- Farazouli A, Cerratto-Pargman T, Bolander-Laksov K & McGrath C 2024. Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, 49(3):363–375. <https://doi.org/10.1080/02602938.2023.2241676>
- Farrokhnia M, Banihashem SK, Noroozi O & Wals A 2024. A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3):460–474. <https://doi.org/10.1080/14703297.2023.2195846>
- Friederichs H, Friederichs WJ & März M 2023. ChatGPT in medical school: How successful is AI in progress testing? *Medical Education Online*, 28(1):2220920. <https://doi.org/10.1080/10872981.2023.2220920>
- Huang Y, Gomaa A, Semrau S, Haderlein M, Lettmaier S, Weissmann T, Grigo J, Tkhatayt HB, Frey B, Gaipf U, Distel L, Maier A, Fietkau R, Bert C & Putz F 2023. Benchmarking ChatGPT-4 on a radiation oncology in-training exam and Red Journal Gray Zone cases: Potentials and challenges for ai-assisted medical education and decision making in radiation oncology. *Frontiers in Oncology*, 13:1265024. <https://doi.org/10.3389/fonc.2023.1265024>
- Huh S 2023. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: A descriptive study. *Journal of Educational Evaluation for Health Professions*, 20:1. <https://doi.org/10.3352/jeehp.2023.20.01>
- Javier DRC & Moorhouse BL 2023. Developing secondary school English language learners' productive and critical use of ChatGPT. *TESOL Journal*, 15(2):e755. <https://doi.org/10.1002/tesj.755>
- Karabacak M, Ozkara BB, Margetis K, Wintermark M & Bisdas S 2023. The advent of generative language models in medical education. *JMIR Medical Education*, 9:e48163. <https://doi.org/10.2196/48163>
- Kelly A, Sullivan M & Strampel K 2023. Generative artificial intelligence: University student awareness, experience, and confidence in use across disciplines. *Journal of University Teaching and Learning Practice*, 20(6):1–16. <https://doi.org/10.53761/1.20.6.12>
- Khosravi T, Al Sudani ZM & Oladnabi M 2023. To what extent does ChatGPT understand genetics? *Innovations in Education and Teaching International*:1–10. <https://doi.org/10.1080/14703297.2023.2258842>
- Kortemeyer G 2023. Could an artificial-intelligence agent pass an introductory physics course? *Physical Review Physics Education Research*, 19:010132. <https://doi.org/10.1103/PhysRevPhysEducRes.19.010132>
- Lahat A, Shachar E, Avidan B, Glicksberg B & Klang E 2023. Evaluating the utility of a large language model in answering common patients' gastrointestinal health-related questions: Are we there yet? *Diagnostics*, 13(11):1950. <https://doi.org/10.3390/diagnostics13111950>
- Lozano A & Fontao CB 2023. Is the education system prepared for the irruption of artificial intelligence? A study on the perceptions of students of primary education degree from a dual perspective: Current pupils and future teachers. *Education Sciences*, 13(7):733. <https://doi.org/10.3390/educsci13070733>
- Michel-Villarreal R, Vilalta-Perdomo E, Salinas Navarro DE, Thierry-Aguilera R & Gerardou FS 2023. Challenges and opportunities of generative AI for



- higher education as explained by ChatGPT. *Education Sciences*, 13(9):856. <https://doi.org/10.3390/educsci13090856>
- Morjaria L, Burns L, Bracken K, Ngo QN, Lee M, Levinson AJ, Smith J, Thompson P & Sibbald M 2023. Examining the threat of ChatGPT to the validity of short answer assessments in an undergraduate medical program. *Journal of Medical Education and Curricular Development*, 10:1–7. <https://doi.org/10.1177/23821205231204178>
- Netto NR 2023. Use of case studies in social work assessments – ChatGPT’s kryptonite? *Social Work Education*:1–12. <https://doi.org/10.1080/02615479.2023.2266461>
- OpenAI 2023. *GPT-4 technical report*. [Preprint]. <https://doi.org/10.48550/arXiv.2303.08774>
- OpenAI n.d. *Advice and answers from the OpenAI Team*. Available at <https://help.openai.com/en/>. Accessed 23 November 2023.
- Ouh EL, Gan BKS, Shim KJ & Wlodkowski S 2023. ChatGPT, can you generate solutions for my coding exercises? An evaluation on its effectiveness in an undergraduate Java programming course. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*. New York, NY: ACM. <https://doi.org/10.1145/3587102.3588794>
- Preferred Reporting Items for Systematic reviews and Meta-Analyses 2023. *PRISMA transparent reporting of systematic reviews and meta-analyses*.
- Shaikh S, Yayilgan SY, Klimova B & Pikhart M 2023. Assessing the usability of ChatGPT for formal English language learning [Special issue]. *European Journal of Investigation in Health, Psychology and Education*, 13(9):1937–1960. <https://doi.org/10.3390/ejihpe13090140>
- Shue E, Liu L, Li B, Feng Z, Li X & Hu G 2023. *Empowering beginners in bioinformatics with ChatGPT*. [Preprint]. <https://doi.org/10.1101/2023.03.07.531414>
- Singh H, Tayarani-Najaran MH & Yaqoob M 2023. Exploring computer science students’ perception of ChatGPT in higher education: A descriptive and correlation study. *Education Sciences*, 13(9):924. <https://doi.org/10.3390/educsci13090924>
- Singh M 2023. Maintaining the integrity of the South African university: The impact of ChatGPT on plagiarism and scholarly writing. *South African Journal of Higher Education*, 37(5):203–220. <https://doi.org/10.20853/37-5-5941>
- Snyder H 2019. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104:333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Stojanov A 2023. Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20:35. <https://doi.org/10.1186/s41239-023-00404-7>
- Su J & Yang W 2023. Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, 6(3):355–366. <https://doi.org/10.1177/20965311231168423>
- Tawfik GM, Dila KAS, Mohamed MYF, Tam DNH, Kien ND, Ahmed AM & Huy NT 2019. A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Tropical Medicine and Health*, 47(1):46. <https://doi.org/10.1186/s41182-019-0165-6>
- Tirado-Olivares S, Navío-Inglés M, O’Connor-Jiménez P & Cózar-Gutiérrez R 2023. From human to machine: Investigating the effectiveness of the conversational AI ChatGPT in historical thinking. *Education Sciences*, 13(8):803. <https://doi.org/10.3390/educsci13080803>
- Tülübaş T, Demirkol M, Ozdemir TY, Polat H, Karakose T & Yirci R 2023. An interview with ChatGPT on emergency remote teaching: A comparative analysis based on human–AI collaboration. *Educational Process: International Journal*, 12(2):93–110. <https://doi.org/10.22521/edupij.2023.122.6>
- Ullah R, Ismail HB, Khan MTI & Zeb A 2024. Nexus between Chat GPT usage dimensions and investment decisions making in Pakistan: Moderating role of financial literacy. *Technology in Society*, 76:102454. <https://doi.org/10.1016/j.techsoc.2024.102454>
- Waltzer T, Cox RL & Heyman GD 2023. Testing the ability of teachers and students to differentiate between essays generated by ChatGPT and high school students. *Human Behavior and Emerging Technologies*, 2023:1923981. <https://doi.org/10.1155/2023/1923981>
- Wang K 2024. From ELIZA to ChatGPT: A brief history of chatbots and their evolution. *Applied and Computational Engineering*, 39:57–62. <https://doi.org/10.54254/2755-2721/39/20230579>
- Yang C & Stivers A 2024. Investigating AI languages’ ability to solve undergraduate finance problems. *Journal of Education for Business*, 99(1):44–51. <https://doi.org/10.1080/08832323.2023.2253963>
- Zeb A, Ullah R & Karim R 2024. Exploring the role of ChatGPT in higher education: Opportunities, challenges and ethical considerations. *International Journal of Information and Learning Technology*, 41(1):99–111. <https://doi.org/10.1108/IJILT-04-2023-0046>
- Zhu C, Sun M, Luo J, Li T & Wang M 2023. How to harness the potential of ChatGPT in education? *Knowledge Management & E-Learning*, 15(2):133–152. <https://doi.org/10.34105/j.kmel.2023.15.008>