

Research

Exploring the difference and quality of AI-generated versus human-written texts

Karin Tengler¹ · Gerhard Brandhofer¹

Received: 4 June 2024 / Accepted: 5 May 2025

Published online: 13 May 2025

© The Author(s) 2025 OPEN

Abstract

Generative Artificial Intelligence (GenAI) models have grown increasingly popular among pre-service teachers (PSTs) and have become their constant companions, primarily assisting them in scientific writing. This article presents a study that investigates the differences and benefits of GenAI in the scientific writing process. Essays generated by GenAI were compared with those written by pre-service teachers ($N=39$). The results of the criticism-led evaluation demonstrate that AI-generated texts perform significantly better, particularly in logic and fluency. In addition, most PSTs perceive generative AI beneficial in various phases of the writing process. The findings highlight the potential of GenAI as a useful tool in higher education, but also raise questions regarding academic integrity and critical thinking.

Keywords Artificial intelligence · Higher education · Generative AI · Writing process

1 Introduction

Generative Artificial Intelligence (GenAI) passes tests and has quickly become a constant companion for students. According to a German study, 98% of students have already heard of generative AI, and 78% use it for scientific and private purposes [1]. Current acceptance studies, such as the one by Brandhofer and Tengler [2], also show that not only students but also teachers and pre-service teachers (PSTs) are generally very optimistic about the use of GenAI. Alongside Google Gemini and Microsoft Co-Pilot, ChatGPT from OpenAI is currently one of the best-known AI tools that can generate human-like texts, among other things [3]. The rapid spread and easy access to AI-supported language models quickly made it clear that they would not only have a far-reaching impact on numerous areas of life but would also bring about significant changes in the education system, which is once again facing significant change following the COVID-19 pandemic [4]. AI is not only found in text and image generators but also in technologies such as automatic speech and facial recognition, personalized social media feeds, individually tailored music, and purchase recommendations [5]. AI text and image generators are often accessible free of charge in their basic version. They are characterized by high performance and user-friendliness and have thus become firmly established in society [6]. Generative AI like ChatGPT can enhance the learning and creativity of pre-service teachers by facilitating multi-round conversations on various topics, thus improving the development of teaching aids [7]. AI-supported applications are used to answer questions, solve problems, explain complex topics, provide virtual tutoring, practice languages and write essays [3]. The potential of generative AI makes it possible to generate texts almost indistinguishable from those written by humans. Therefore, this article presents a study investigating to what extent AI-generated texts are recognized as such and to what extent they differ qualitatively from texts written by humans.

✉ Karin Tengler, k.tengler@ph-noe.ac.at | ¹University College of Teacher Education Lower Austria, Baden, Austria.



2 Theoretical background

2.1 Use of generative AI in the writing process

The technologies currently in use are based on self-learning language models, so-called Large Language Models (LLMs), which can continuously expand their pattern recognition capabilities through training with extensive text data. User queries are answered so effectively that the generated texts can create all text types. GenAI offers a variety of examinations on a range of topics, and LLMs serve as a great writing assistance [8]. Rahman and Watanobe [3] describe ChatGPT as "a revolutionary LLM that can maintain human-like conversations and generate human-like text for any natural language query that is nearly indistinguishable". In addition to the many benefits that generative AI brings to studies, it also poses challenges, especially in education. As text generators can produce scientific texts on all kinds of topics that appear human-like, teachers and educational institutions also need to be aware of potential cheating risks in written work and exams. Likewise, Schirmer et al. [5] emphasize that implementing AI tools in educational contexts is an inevitable development that offers a variety of pedagogical and methodological challenges and opportunities. The effective use of generative AI technologies in education requires the development and implementation of specific formats as well as in education, training and further education [9]. Stolpe and Hallström [10] expand on this discussion by describing the importance of conceptual, procedural, and contextual knowledge for a comprehensive understanding of AI in education. The most fundamental support text generators provide when writing texts is at the syntactic level, for example, by recognizing and correcting typos. Semantically, these models can identify grammatical inconsistencies and point out possible improvement strategies. In addition, they offer the possibility to realize stylistic improvements on a topic-specific basis [9, 11]. For over two decades, correcting spelling and grammatical errors in word processing programs has been automated using simple databases that match words and sentences, sometimes with errors. In addition to the easy access and simplified handling of AI text generators, the use of generative AI has also led to a significant increase in the quality of autocorrection results for some time now [6].

The great potential of LLMs lies in the ability to understand and produce (written) natural language [5]. Consequently, technologies such as ChatGPT prove to be valuable tools in the writing process, for example when writing essays [12]. The writing process itself is complex, dynamic, and non-linear. Writing research identifies three sub-processes: Planning, writing and revising [13]. Gen AI has the potential to assist in all phases of the writing process, including the generation of ideas and arguments, the organization of content, the correction of linguistic errors, and the suggestion of alternative sentence structures to enhance coherence and stylistic quality [12]. Didactic support in writing and reflection on these processes are essential to meaningfully evaluate the use of AI tools and decide when their use is appropriate [6].

2.2 Related work

With the rapid increase in the use of generative AI to support the writing process, this has also increasingly become the subject of research. Von Garrel et al. [14] mention in their publication various practical applications of AI-based tools in the writing process, including conducting literature searches, translating into multiple languages, clarifying comprehension issues, and developing concepts. Moreover, generative AI can draw attention to underexplored dimensions and contemporary research topics, thereby facilitating a deeper comprehension and analysis of complex issues among university students. Through its reflective use, it may also contribute to the cultivation of critical thinking and problem-solving abilities [11].

Chan and Hu [15] investigated university students' views on generative AI in higher education. Students expressed a generally positive attitude toward technologies such as ChatGPT and valued its support in academic writing. This perspective is also reflected in the findings of Malik et al. [12] in their study on academic essay writing, who report on the potential benefits of generative AI in promoting writing creativity and facilitating the creation of original artworks and literature. In their study, Hakam et al. [16] examine the capabilities of AI-generated texts in comparison to scientific texts written by humans. The study confirms that AI tools are capable of producing structurally sound and coherent texts; however, these texts often lack depth and critical analysis. AI-generated content tends to rely on generalizations or superficial statements rather than offering original insights or substantively grounded arguments. Overall, generative AI currently serves more as a complement than a replacement for human intellect and creativity, highlighting the continued need for human oversight to ensure originality and scientific rigor. Amirjalili et al. [17] aimed to recognize AI-generated academic essays and identify the necessary competencies to do so, showing that instructors often resort to error-oriented

assessment methods. In these methods technical and grammatical perfection rather than evidence of AI work is seen in assessment, and checking the accuracy of facts and proper source citation was neglected. Furthermore, this study underscores the importance of reviewing and strengthening institutional policies and practices that protect academic integrity and ensure quality higher education. In their study, Muñoz-Ortiz et al. [18] focused on identifying linguistic patterns between human-written English news texts and outputs from six different large language models (LLMs) by quantitatively analyzing and comparing them. The study explores several linguistic dimensions, including morphological, syntactic, psychometric and sociolinguistic aspects. The results demonstrate significant differences between the human texts and the LLMs. The human texts display greater variability in sentence length and vocabulary, a different use of grammatical structures and a greater tendency to express stronger negative emotions. In contrast, the LLM output shows a greater use of numbers and symbols, indicating a more objective language style [18]. Additionally, it displays a sexism bias that is present in human texts but amplified in machine output. Basic et al. [19] investigated the extent to which the use of ChatGPT-3 influences students' scientific writing in a group of Master students. The research focused on the possible effects of text authenticity change on grade improvement, writing time reduction, and text authenticity. Although ChatGPT can generate contextually relevant content, the results show that the use of ChatGPT-3 did not lead to a significant increase in essay quality. The group that used ChatGPT-3 for support did not achieve better results in terms of text content and speed, but the use had positive effects on idea generation and to speed up the initial writing phase. One possible explanation was that the students had not yet been familiar with the tool. In their study, Lindqvist and Arvidsson [20] examine the function of reflection as a pedagogical instrument and contrast genuine student reflection texts with those generated by AI (ChatGPT). The findings indicate that the student texts encompass personal insights and methodologies to facilitate learning, whereas the AI texts predominantly present comprehensive reflection questions that foster critical discourse. The AI-generated texts did not pose a threat to academic integrity but rather offered a valuable adjunct to assist students in intensifying their reflections. The study underscores the potential benefits of AI in education, particularly in the promotion of reflection and learning [20].

Herbold et al. [21] compared essays written by high school students with those generated by ChatGPT-3 and ChatGPT-4 in their study. Each of the 139 participating teachers rated six texts according to various criteria, such as topic and completeness. The study shows that ChatGPT can generate high-quality argumentative essays and significantly exceeds the quality of human-written essays in all evaluation categories. It was also evident that ChatGPT-4 also delivered even better results than its predecessor. The writing style of GenAI differs in that ChatGPT models use more nominalizations and the sentences have a more complex sentence construction.

The ability of generative AI models to produce semantically and syntactically correct texts is also leading to a profound change in how scientific texts are created and evaluated in higher education. This raises important research questions. This study aims to contribute to research by answering the following research questions.

- To what extent is it possible to recognize whether a text was written by a human or generative AI?
- To what extent do AI-generated texts differ from human-written texts regarding language and content?
- In which aspects of the writing process do PSTs perceive generative AI to be beneficial?

3 Methodology

3.1 Participants and setting

In 2023/24, the Bachelor degree course in Primary Education included the teaching session Children learn digitally (*teaching and learning with digital media, the design of digital learning spaces*) as part of the subject digitization, which focuses on implementing digital media and discussion of STEM topics, consists of a lecture and a workshop. The content and structure of the course are derived from the digi.kompP, a competence model for digital competencies for educators [22]. First-semester students were required to enroll in this course.

In the lecture, the pre-service teachers were introduced to the topic of artificial intelligence, focusing on the basics of GenAI and its didactic use. As part of the associated workshop, the students, who were divided into two groups in the workshop, were each assigned to write an essay of 200 words on the topic of *Digital media in primary education—potentials and risks* and to generate two essays on the same length by various AI tools such as Chat GPT, Google Gemini and Microsoft Co-Pilot. The prompt to be entered was: Write an essay of around 200 words on *Digital media in primary education—potentials and risks*. The texts were entered into a form by the PSTs, uploaded to the learning platform and

anonymized by the research team. The PSTs in the other group then received two randomly selected texts, which were initially assessed as AI-generated or not and evaluated according to predefined criteria. This assessment was entered into an online questionnaire.

A total of 39 student teachers participated in creating and evaluating the essays. This meant that 78 texts, 41 AI-generated texts and 37 human-written texts, could be evaluated. Submitting the texts also served as consent to participate in the study. The authors did not consider it relevant to classify the participants by gender for this study.

3.2 Questionnaire

The questionnaire, created using Microsoft Forms, consists of three sections. The first section contains seven categories, which were used based on the study by Herbold et al. [21]. These include topic and completeness, logic and composition, expressiveness and comprehensiveness, language mastery, complexity, vocabulary and text linking, and language constructs. For each category, a seven-point Likert scale was established with scores ranging from zero to six, with zero being the worst score (e.g., no logical structure) and six being the best score (e.g., clear structure, logical and sound argumentation). Students were given a list of the categories and a description of the associated rating levels from zero to seven as a guide. The second section deals with the assessment of whether a human or a generative AI created the text. Finally, the PSTs were asked about their attitudes towards the usefulness of AI-supported text generators in the writing process. The proposed items were taken from the studies mentioned in the theory section [1, 14]. Each item could be rated on a four-point Likert scale from 4 (agree) to 1 (disagree).

3.3 Data analysis

As all questionnaires were duly completed, all data records were exported from Microsoft Forms and could therefore be analyzed. The statistical analysis was done using IBM SPSS 29 software, starting with the mean value. To identify potential differences in the evaluation of the texts with regard to linguistic and content-related categories, a paired-samples t-test was applied [23]. The paired t-test is used to statistically analyze whether the means of two related samples are significantly different. In addition, Cohen's d effect size is used, which analyses the size of the difference between two means expressed in standard deviations. Hedge's g correction is used to provide a more accurate estimate of effect size for small samples. The data regarding the perceived benefits of GenAI in the writing process were analyzed using descriptive statistics. The results are presented and discussed in the following section.

4 Results

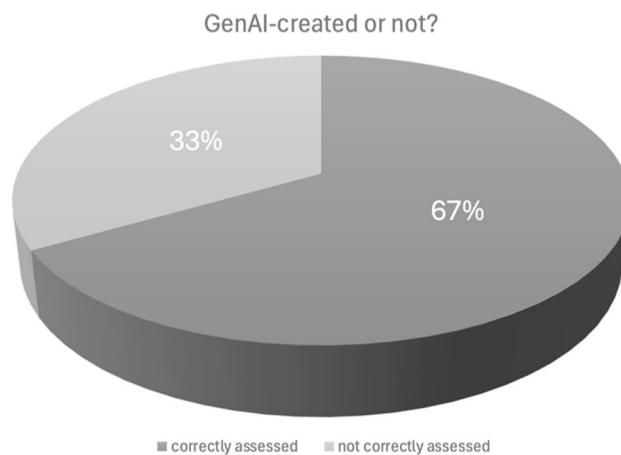
The results are presented in three stages. First, students' assessments of whether a text was written by GenAI or a human are introduced. This is followed by an analysis of the evaluations of linguistic and content-related aspects for both groups of texts. Finally, students' perceptions of the usefulness of generative AI across different phases of the writing process are presented.

4.1 AI or not?

The pie chart (see Fig. 1) shows the results of the PSTs' assessment of whether the essays submitted for evaluation were written with or without the help of artificial intelligence.

52 of the 78 texts were correctly assessed by the PSTs as to whether or not a text generator had been used to produce the essay. This means that the majority, exactly two-thirds of the assessments (66.67%), were correct. This could indicate that, in many cases, AI-generated texts are not easy to distinguish from human-generated texts or that students are not yet fully able to recognize the differences.

Fig. 1 Assessment, created by AI or not?



4.2 Text analysis

The research question aimed to determine the extent to which AI-generated texts differ from human-written texts regarding language and content. Table 1 shows the statistical results after applying a paired t-test, used in empirical research to identify significant differences between dependent or independent samples [23]. In this case, the aim was to analyze the students' evaluation of the texts in the categories of topic and completeness, logic and composition, expressiveness and comprehensibility, command of language, complexity of sentence structures, vocabulary and text linking, and language construction.

The effect sizes (Cohen's d and Hedge's g), which indicate how significant the difference between the groups is [23], are in the range of small to medium effect for most pairs. The Hedge's g correction is used to obtain a more accurate estimate of the effect size for small samples.

Topic and completeness The small effect size (Cohen's $d = 0.071$) and the 95% confidence interval indicate no statistically significant difference in topic treatment and completeness between AI-generated and human-written texts. The small difference in topic coverage and completeness between AI-generated and human-authored texts shows that AI tools can effectively cover the assigned topic.

Logic and composition The mean effect size (Cohen's $d = 0.317$) and the confidence interval indicate that AI-generated texts perform better in terms of logical structure than texts written by humans. This is a beneficial and advantageous feature of AI-generated texts, particularly for the composition of functional texts. It demonstrates that AI tools are not only capable of producing content that is relevant and pertinent, but also of presenting it in a logical and coherent manner.

Expressiveness and comprehensiveness Due to the small effect size (Cohen's $d = 0.104$), no significant difference in the expressiveness and comprehensibility of both text groups is recognizable here. In order to more accurately identify potential significant differences in the category of Expressiveness and comprehensiveness, it would have been preferable to formulate the topic more precisely.

Mastery of the language The moderate effect size (Cohen's $d = 0.355$) and the confidence interval indicate that AI-generated texts show a better mastery of the language, possibly due to the correct use of grammar and syntax, which is one of the great advantages of generative AI.

Complexity Despite the moderate effect size (Cohen's $d = 0.237$), the confidence interval is not meaningful enough to confirm a significant difference in the complexity of the texts. One potential explanation for this outcome is the relatively brief nature of the task, which required a limited number of words.

Vocabulary and text linking Here, too, the moderate effect size (Cohen's $d = 0.244$) with a confidence interval that includes zero does not allow a clear conclusion to be drawn about significant differences in vocabulary and text linking between the two text groups. This result suggests that while AI can deliver adequate content, the depth and more sophisticated vocabulary required in higher education or advanced discussions still needs to be developed for AI texts to perform significantly better in these contexts.

Language construct The small effect size (Cohen's $d = 0.117$) indicates that no significant differences were found in the language constructs used between AI-generated and human-written texts. One potential explanation is that

Table 1 Data analysis

Category	Mean	Std. Dev	95% Conf.I./ DiffLower	95% Conf.I./ DiffUpper	t	df	One-Sided p	Two-Sided p	Cohen's d	Hedges' corr
Topic and completeness	-0.081	1.140	-0.299	0.461	0.433	36	0.334	0.668	0.071	0.070
Logic and composition	0.405	1.279	-0.021	0.832	1.928	36	0.031	0.062	0.317	0.310
Expressiveness/comprehensiveness	0.162	1.555	-0.356	0.681	0.634	36	0.265	0.530	0.104	0.102
Language mastery	0.514	1.446	0.031	0.996	2.161	36	0.019	0.37	0.355	0.348
Complexity	0.486	2.050	-0.197	1.170	1.444	36	0.079	0.157	0.237	0.232
Vocabulary/text linking	0.432	1.772	-0.158	1.023	1.484	36	0.073	0.146	0.244	0.239
Language constructs	0.189	1.613	-0.349	0.727	0.713	36	0.240	0.480	0.117	0.115

text generators are trained with extensive data sets of human texts, which often encompass a diverse range of genres and styles. Consequently, the AI learns to replicate language patterns and structures that are characteristic of human writing. This results in texts generated by the AI exhibiting a high degree of similarity to human texts in terms of their linguistic constructs.

Overall, the results show significant differences in the category of logic and composition as well as in language proficiency, with AI-generated texts generally performing better. The differences are insignificant in other areas, such as expressiveness, complexity, and language construction. The fact that significant differences were only evident in the two categories may also be due to the small sample size. The results suggest that the quality of AI-generated texts is higher than that of texts written by humans in terms of content and linguistic aspects. This underlines the study results by [21], in which the texts generated by ChatGPT-3 and ChatGPT-4 surpassed the quality of human-generated texts.

4.3 Use of text generators

In the last section of the questionnaire, the PSTs were asked in which phases of the writing process they perceived using generative AI to be beneficial. The answers were recorded using a four-point Likert scale.

Table 2 provides insights into students' perceptions of how useful generative AI is as a support in the various phases of the writing process. The evaluation and interpretation of the assessment can be found below.

Writing texts With an average value of $M = 2.92$ ($SD = 0.900$), students rate the usefulness of AI for basic text writing as moderate. The high standard deviation indicates heterogeneous opinions, which suggests that the perceived usefulness of generative AI in essay writing is viewed very controversially from person to person.

Structuring and concept creation Text generators are perceived as quite applicable as a tool for structuring and creating concepts ($M = 3.38$, $SD = 0.633$). This also reflects the results of AI studies [1, 14, 16].

Finding arguments With the highest average rating ($M = 3.54$, $SD = 0.643$), students seem to find generative AI particularly effective in generating arguments for their texts. The low standard deviation indicates that the students are largely in agreement here. This is also consistent with the results of the studies by Basic et al. [19] and Malik et al. [12].

Precise formulation Students rated the ability of text generators to help with precise wording as less useful overall ($M = 2.67$, $SD = 0.838$). However, the higher standard deviation and a look at the diagram show that the answers are not very homogeneous, as almost half still perceived AI as a positive tool for accurate formulation.

Overcoming writer's block Here, too, the potential benefits of AI are acknowledged ($M = 3.38$, $SD = 0.847$). Although the range of ratings and the standard deviation indicate different experiences, most students still find that text generators can be seen as very helpful in overcoming writer's block. This aligns with the perceptions in the study by Basic et al. [19].

Error prevention Although this area scored the lowest in this survey ($M = 2.54$, $SD = 0.884$), more than half of the respondents gave it a positive rating. This may be due to the wording of the item. Perhaps the term *error correction* would have been more appropriate.

The assessments of the perceived benefits conclude that text generators are predominantly seen as positive support in all phases of the writing process. PSTs primarily use GenAI as a tool for idea development and argumentation, whereby the potential for conceptualizing and structuring texts can be emphasized. In contrast, opinions on the effectiveness of AI in precise formulation and error avoidance are divided, which could indicate limitations in these areas. Overall, the evaluation shows that generative AI is seen as a very useful and supportive tool in the writing process.

Table 2 Perceived benefit/
descriptive statistics

Category	N	Mean	Std. Deviation	Minimum	Maximum
Writing texts	39	2.92	0.900	1	4
Structuring and concept creation	39	3.38	0.633	2	4
Finding arguments	39	3.54	0.643	2	4
Precise formulation	39	2.67	0.838	1	4
Overcoming writer's block	39	3.38	0.847	1	4
Error prevention	39	2.54	0.884	1	4

5 Discussion

The significant differences in logic and completeness between AI-generated and human-written texts, as evidenced in the study, are consistent with the findings presented in the theoretical background. In accordance with the theoretical framework, LLMs such as ChatGPT offer substantial assistance in the writing process, particularly regarding text structuring and grammatical correctness. The evidence suggests that AI-generated texts tend to perform better, although the improvements are not consistently significant across all areas. This is in line with the findings of previous studies, such as those by Lindqvist and Arvidsson [20]. Additionally, the study revealed that AI-generated texts predominantly included comprehensive reflection questions and provided responses to them. The study by Hakam et al. [16], which emphasizes the ability of AI tools to produce structurally coherent texts, also corroborates our findings, demonstrating a superior logical structure in AI-generated texts.

Additionally, our findings indicate that AI-generated texts demonstrate superior fluency, as evidenced by the moderate effect size observed in this category. This aligns with the observations of Kasneci et al. [11], who highlight that AI text generators produce accurate texts in terms of grammar and spelling, enabling them to identify inconsistencies and suggest potential improvement strategies.

With regard to the expressiveness and comprehensiveness of texts, our findings indicate no statistically significant differences between AI-generated and human-authored texts. This suggests that, in their current form, AI tools may not be capable of fully capturing or replicating the depth of content observed in human texts. This may be an area where human writers retain an advantage, as evidenced by the findings of Muñoz-Ortiz et al. [18], which indicate that human texts exhibit greater variability in sentence length and vocabulary, as well as a tendency towards more negative wording. This fact was not noticeable in the present study, likely due to the relatively brief length of the texts. The research by Malik et al. [12] investigating the potential of AI in academic text creation and the results indicating that AI tools can be beneficial in promoting writing creativity and facilitating the creation of original artworks and literature support our findings on the usefulness of AI in language mastery and logical structure.

The results of the study comparing AI-generated and human-written texts have significant implications for academic integrity and educational practice. While AI tools have the potential to facilitate the writing process by providing structural and grammatical support, they could also pose a risk to academic honesty if students use them to complete assignments that typically require independent creative or analytical thinking and potentially generate plagiarized academic work. Such an approach could potentially compromise the integrity of student work, necessitating the development of new guidelines and modifications to examination formats.

In addition, reliance on AI could undermine the development of independent writing and thinking skills. Students who become overly reliant on technology risk neglecting crucial aspects of the learning process, which could have a long-term impact on the quality of education. Educational institutions therefore face the challenge of adapting curricula to ensure that AI tools enhance learning without undermining critical thinking and individual analysis skills. On the other hand, AI tools could promote critical thinking by speeding up access to information and allowing more time for deeper analysis and evaluation of content. Educators could use these technologies to enrich teaching and support personalized learning but need to ensure responsible use and avoid encouraging over-reliance on technology.

One potential limitation of the study is the relatively small sample size, which may reduce the statistical power of the analysis and increase the risk of sample bias. Additionally, the fact that the writing task was relatively short may be insufficient to fully capture the complexity of the text. Furthermore, it is possible that student ratings may be influenced by personal biases or an inability to recognize complex textual nuances. Relying solely on student ratings, without incorporating evaluations by academic experts, may affect the reliability and validity of the results. However, it should be considered that the students assessing the texts were pre-service teachers, whose academic training likely provided them with a relatively high level of evaluative competence. Still, future studies should aim to include a larger and more diverse sample, as well as expert ratings, to strengthen the accuracy and generalizability of the findings.

6 Conclusion and outlook

This study illustrates the growing potential and increasing acceptance of generative AI systems in education, particularly in the context of the writing process. The results show that AI-generated texts can perform significantly better than human-written texts in certain aspects of the writing process. The PSTs' assessment of the usefulness of GenAI for writing processes documents that it offers valuable support, especially in the generation of ideas and argumentation, but is also seen as a valuable tool in other phases of writing, especially scientific papers.

Considering these findings, educational institutions should continue to promote the integration of AI tools, but with a critical eye on maintaining academic integrity and developing critical thinking. It is also crucial that PSTs, in particular, are trained in the use of GenAI and fully informed about its possibilities and risks. Future research could aim to develop specific pedagogical strategies and guidelines that make the use of AI in higher education practical and ethical.

Given the rapid development and diverse applications of AI tools, it is essential that teachers and students acquire both the technical skills and critical awareness to use this technology. This will improve the quality and reputation of academic work and ensure informed and reflective use of digital technologies in education.

Author contributions K.T. and G.B. are responsible for the research and data analysis, both wrote the manuscript. Both authors reviewed the manuscript.

Funding The researchers funded this study independently, without external financial support.

Data availability The authors confirm that all data generated or analysed during this study are included in this manuscript.

Declarations

Ethics approval and consent to participate The protocol was approved by the University College of Education Lower Austria and is in accordance with the ethical standards set out by the Austrian Agency for Research Integrity (ÖAWI).

Informed consent All participants in the study provided informed consent for their participation and for the publication of the study's findings.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Schlude A, Mendel U, Stürz R, Fischer M. Verbreitung und Akzeptanz generativer KI an Schulen und Hochschulen. Bidt. 2024. <https://www.bidt.digital/publikation/verbreitung-und-akzeptanz-generativer-ki-an-schulen-und-hochschulen/>
2. Brandhofer G, Tengler K. Zur Akzeptanz von KI-Applikationen bei Lehrenden und Lehramtsstudierenden. R&E-Source. 2024;11(3):7–25. <https://doi.org/10.53345/resource.2024.i3.a1277>.
3. Rahman MM, Watanobe Y. ChatGPT for education and research: opportunities, threats, and strategies. Appl Sci. 2023;13(9):5783. <https://doi.org/10.3390/app13095783>.
4. Albrecht S. ChatGPT und andere Computermodelle zur Sprachverarbeitung–Grundlagen, Anwendungspotenziale und mögliche Auswirkungen. Karlsruher Institut für Technologie. 2023. <https://doi.org/10.5445/IR/1000158070>
5. Schirmer K, Berger M, Himpisl-Gutermann K, Lorenz S-A, Steiner M. Künstliche Intelligenz im Unterricht: Lehr-/Lernszenarien für verschiedene Gegenstände. Medienimpulse. 2023. <https://doi.org/10.21243/mi-02-23-07>.
6. Wampfler P. Schreiben mit KI-Tools. 2023. 2023/3. <https://doi.org/10.58098/lffi/2023/03/806>
7. Ji Y, Zou X, Li T, Zhan Z. The effectiveness of ChatGPT on pre-service teachers' STEM teaching literacy, learning performance, and cognitive load in a teacher training course. Proceedings of the 2023 6th International Conference on Educational Technology Management, 2023; 16–22. <https://doi.org/10.1145/3637907.3637948>

8. Jahic I, Ebner M, Schön S. Harnessing the power of artificial intelligence and ChatGPT in education—a first rapid literature review. Proceedings of EdMedia+ Innovate Learning, 2023; 1462–1470.
9. Jeon J, Lee S. Large language models in education: a focus on the complementary relationship between human teachers and ChatGPT. *Educ Inf Technol.* 2023;28(12):15873–92. <https://doi.org/10.1007/s10639-023-11834-1>.
10. Stolpe K, Hallström J. Artificial intelligence literacy for technology education. *Comput Educ Open.* 2024;6:100159. <https://doi.org/10.1016/j.caeo.2024.100159>.
11. Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Gasser U, Groh G, Günemann S, Hüllermeier E. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn Individ Differ.* 2023;103:102274. <https://doi.org/10.1016/j.lindif.2023.102274>.
12. Malik AR, Pratiwi Y, Andajani K, Numertayasa IW, Suharti S, Darwis A. Exploring artificial intelligence in academic essay: higher education student's perspective. *Int J Educ Res Open.* 2023;5:100296. <https://doi.org/10.1016/j.ijedro.2023.100296>.
13. Sieberer E. Schreibkompetenz vermitteln: Ein Blick auf die Unterrichtspraxis. R&E-SOURCE. 2020. <https://journal.ph-noe.ac.at/index.php/resource/article/view/915>
14. von Garrel J, Mayer J, Mühlfeld M. Künstliche Intelligenz im Studium Eine quantitative Befragung von Studierenden zur Nutzung von ChatGPT & Co. 2023. <https://doi.org/10.1057/s41599-023-02304-7>
15. Chan CKY, Hu W. Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *Int J Educ Technol High Educ.* 2023;20(1):43. <https://doi.org/10.1186/s41239-023-00411-8>.
16. Hakam HT, Prill R, Korte L, Lovreković B, Ostožić M, Ramadanov N, Muehlensiepen F. Human-written vs AI-generated texts in orthopedic academic literature: comparative qualitative analysis. *JMIR Formative Res.* 2024;8: e52164. <https://doi.org/10.2196/52164>.
17. Amirjalili F, Neysani M, Nikbakht A. Exploring the boundaries of authorship: a comparative analysis of AI-generated text and human academic writing in English literature. *Front Educ.* 2024. <https://doi.org/10.3389/feduc.2024.1347421>.
18. Muñoz-Ortiz A, Gómez-Rodríguez C, Vilares D. Contrasting linguistic patterns in human and LLM-generated news text. *Artif Intell Rev.* 2024. <https://doi.org/10.1007/s10462-024-10903-2>.
19. Basic Z, Banovac A, Krizic I, Jerkovic I. Better by you, better than me, chatgpt3 as writing assistance in students essays. arXiv preprint. 2023. <https://doi.org/10.48550/arXiv.2302.04536>
20. Lindqvist MH, Arvidsson C. Exploring student and AI generated texts: reflections on reflection texts. *Electron J e-Learn.* 2024;22(6):52–9. <https://doi.org/10.34190/ejel.22.6.3473>.
21. Herbold S, Hautli-Janisz A, Heuer U, Kikteva Z, Trautsch A. A large-scale comparison of human-written versus ChatGPT-generated essays. *Sci Rep.* 2023;13(1):18617. <https://doi.org/10.1038/s41598-023-45644-9>.
22. Brandhofer G, Miglbauer M, Fikisz W, Höfler E, Kayali F. Die Weiterentwicklung des Kompetenzrasters digi. kompP für Pädagog innen. In: Bildung und Digitalisierung. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG; 2020. p. 51–72. <https://doi.org/10.5771/978348906247-51>.
23. Döring N. Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. Berlin: Springer; 2023. <https://doi.org/10.1007/978-3-662-64762-2>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.