Data Inputs and Contexts for ChatGPT in Education: A Review of Empirical Studies

Wenting Sun[1], Jiangyue Liu[2], Xiaoling Wang[3]

Humboldt-Universität zu Berlin, Germany[1]
Suzhou University, China[2]
Zhejiang Normal University, China[3]

Abstract

Despite some education stakeholders' attempts to ban ChatGPT, many support its meaningful and effective use in teaching and learning [1,2,3]. [4] argues that the misuse of this tool often results from a lack of knowledge and understanding of its proper use and encourages the development of clear guidelines for the acceptable use of AI tools like ChatGPT. Recently, significant efforts have been made to design appropriate prompts. However, the data and context for their use in education remain understudied.

This review, following the Preferred reporting items for systematic reviews and meta-analysis (PRISMA) guidance, selected 102 empirical articles with detailed prompts when use ChatGPT from Web of science and Scopus. A qualitative method was utilized to extract information about educational level, knowledge domain, research method, ChatGPT affordances in education, data source (embedded in prompting), data collection method. Teaching and learning experiences were summarized from these articles.

It was found that higher education is the main educational levels adapting ChatGPT. Language learning and education in humanities, as well as computer science, chemistry, mathematics, and medicine in STEM disciplines, gained more attention. Half of the included articles chose qualitative method. When using and analysing ChatGPT, data collected from or related to course level was the main source for embedding context or task information into prompting, and data generated by ChatGPT was the main data source for further analysis of the ChatGPT outputs based on some prompts (rather than users' perceptions and performance impacted by ChatGPT). Moreover, ChatGPT provides more diverse affordances in two course phases: during the courses and after the courses. Some teaching strategies and learning strategies were summarized to enhance students' critical or reasoning thinking.

Keywords: *ChatGPT in Classroom, Affordances, Prompting, AI Tools in Teaching*

1. Introduction

As a generative artificial intelligence (GenAI) chatbot powered by large language models (LLMs), ChatGPT has demonstrated strengths in the education area. The exploration of integrating ChatGPT in teaching and learning has spread across disciplines in different education levels, such as language education, health professions, engineering, mathematics, computer science, physics, teacher education [1, 4, 5].

Through Socratic dialogues with ChatGPT, students are actively involved in generating ideas, reflecting on learning content, exploring explanations of wrong solutions, and deeply thinking about their solutions [6, 7, 8, 9]. ChatGPT enriches personalized teaching materials and reduces teachers' workload by generating formative and product feedback to improve learner performance [10, 11, 12]. ChatGPT potentially scaffolds younger learners on how to constitute “good” peer feedback in open-ended problem-solving contexts.

ChatGPT exhibits several weaknesses as well. As a chatbot driven by LLMs, ChatGPT is not good at handling images, especially for the free version. The accuracy of ChatGPT’s generated outputs is occasionally questionable [14, 15], which is limited by the training data [16]. The critical and reasoning abilities of ChatGPT do not satisfy users' expectations [5, 17]. It also presents certain challenges to the traditional formal education system. Some misconducts might be facilitated, such as plagiarism, academic dishonesty, productivity, creative and critical thinking decrease hampered by over-reliance on ChatGPT, and superficial learning [1, 4, 5]. The ethical problem is another concern about the usage of ChatGPT, such as data privacy and security [5, 17, 18].

Even though some education stakeholders try to ban the usage of ChatGPT in education, different voices are expressing their support for meaningfully and effectively using ChatGPT to assist teaching and learning [1, 2, 3]. [4] explains that the misuse of this tool often results from a lack of knowledge and understanding of its proper use and encourages the development of clear guidelines for the acceptable use of AI tools like ChatGPT. How to effectively use LLMs-driven ChatGPT is a complex topic, and one of the main concerns is prompting. Prompt engineering has become a focal point in efforts to make ChatGPT more effective in educational contexts. Yet, there remains a gap in understanding the types of data embedded within prompts and how they reflect the teaching and learning contexts in which ChatGPT is deployed. This study aims to synthesize existing empirical research to provide an overview of how ChatGPT is used in education, particularly focusing on the role of contextual data inputs within prompts.

2. Related Works

2.1 Prompting

To effectively and efficiently navigate and enjoy these benefits provided by generative AI tools like ChatGPT, users need some efforts and skills, one of which is prompting [6, 19]. By selecting the appropriate prompts, LLMs can be used to generate the desired output to solve the tasks at hand [20]. Although prompting LLMs appears effortless, designing context sensitive prompt strategies, devising prompts to overcome the arisen error from LLMs, and systematically assessing those prompts strategies' effectiveness is a complex interdisciplinary topic [21]. For the field of education, with the increasing popularity of the prompt engineering practice, investigating the skills of prompt engineering is important [22].

Recently, significant efforts have been made to design appropriate prompts. There are different attempts to categorize prompt construction. For example, depending on the number of examples provided, prompt can differentiate among zero-, one-, and few-shot learning [23]. Few-shot learning employs a strategy in data scarce scenarios to learn a machine learning system with few training samples [20]. However, what data can be part of the prompts and in what usage context in education is still understudied.

2.2 Prior Reviews of ChatGPT in Education

Regarding the reviews of ChatGPT in education, several studies have systematically reported the categories of ChatGPT use in education and analysed usage scenarios. Here, we briefly introduce six reviews to illustrate the research gap we aim to address with our review.

[1] used Biggs's Presage-Process Product (3P) model of teaching and learning to categorize ChatGPT 's strengths, weaknesses, opportunities, and threats (SWOT) in education. They expressed that, comparing to banning the ChatGPT use in short term period, considering how to adapt the curriculum to integrate ChatGPT into education in a meaningful way might be a better long-term strategy for emerging technologies.

[4] summarized the SWOT of the previous systematic review of ChatGPT research. By searching and including empirical studies, the authors examined the influence of ChatGPT on student engagement from behavioural, emotional, and cognitive aspects. They discussed that undesirable behaviours, like plagiarism and cheating, might result from insufficient literacy, knowledge, and understanding of the proper use of ChatGPT.

[5] conducted a systematic review, encompassing the SWOT to analyze ChatGPT's implementation in K-12 educational settings and synthesized ChatGPT usage scenarios from different stakeholders (parents, educators, learners). Prompting engineering was recommended as one concerning in ChatGPT practices.

[17] adopted a three-phased methodology to explore the potentials of ChatGPT in the domain of business education, which consisted of systematic review, semi-structured interviews and text analysis of opinion posts. They observed that ChatGPT usage in pedagogical practice was correlated with students' technical skills, though ChatGPT's function in enhancing skills was complex. The authors outlined several potential research questions about the integration of ChatGPT into education.

Using peer-reviewed studies, [18] looked at ChatGPT's affordances for faculty and students. They also specified the limitations and misuses of ChatGPT. However, their work examined the first eight months following the release of ChatGPT.

[19] reviewed the use of ChatGPT in higher education settings from contextual, methodological, and disciplinary aspects. They discovered that seeking personal benefits and reducing academic burdens were the underlying motivations common across use scenarios.

2.3 This Study

Many reviews defined the diverse categories of ChatGPT usage in education and emphasized integrating ChatGPT into the curriculum. Though prompting engineering significantly impacts ChatGPT output, less is known about how to develop prompts in specific domains or learning tasks, nor how to evaluate the quality of ChatGPT output. It is noted that effective prompting techniques and identification of qualified responses are not only important for ChatGPT usage in education but also for other similar GenAI tools [17].
To contribution to the construction of prompts for ChatGPT in education, this review is guided by two research questions about data embedded in prompting:
RQ1: What general information can be found in empirical articles with prompt details in the field of ChatGPT usage in education (e.g., educational level, knowledge domain, research method, data source embedded in prompting, data collection method)?
RQ2: What teaching and learning experiences can be summarized from the data fed into ChatGPT and their usage context.

3. Methods

3.1 Data Collection

The data were searched in two big comprehensive academic databases, Web of science (WoS) and Scopus. Using topic (title, abstract, keywords) for the search place and English and Chinese for the language, two groups of keywords were used referring to existing reviews we mentioned in Part 2: “chatgpt* OR gpt* OR chatbot* OR Bing OR Bard OR Copilot” AND “learn* OR educat* OR train* OR teach*”. The search procedure followed the Preferred reporting items for systematic reviews and meta-analysis (PRISMA) guidance in the search and selection flow [24]. The research area (education, educational research) was chosen in the WoS database. Peer-reviewed journal papers or long conference papers, or chapters from books were selected as paper type. Initially, 8,622 search results (1,061 from WoS and 7,561 from Scopus) were generated. After abstract scanning and full-text reading, 102 articles were left because they included prompting information in experimental design, research procedure or learning materials parts.

3.2 Data Analysis

We used thematic analysis to extract information about the data used in the selected articles. We followed the six-step proposed by [25]. The extracted data were aggregated in excel for further analysis. For each included articles, we mainly collected three sets of data:
General information: educational level, knowledge domain, research method
Objectives to use ChatGPT: ChatGPT affordances in education,
Data: data source (embedded in prompting), data collection method
Two researchers independently open-coded the articles, with one coding 20% of the data and the other coding all the data. The agreement achieved 85%. The coding results of the selected articles can be found in https://github.com/wentingsunhu/When-Use-ChatGPT-in-Education-What-Data-were-Fed-into-and-during-What-Context-A-Review-of-Empirical

4. Results and Discussion

In general, most studies focused on ChatGPT usage in higher education, with 75 articles (73.5%).

4.1 RQ1

Domains, research methods
Regarding the domains, language learning (17, 16.7%) and computer science (16, 15.7%) gained more attention. The second most popular domains were education (12, 11.7%), while the third most popular were chemistry (9, 8.8%), mathematics (9, 8.8%), and medicine (8, 7.8%). Other domains also

included argumentation, history, physics, management, research methods, 2D animation, and economics. Figure 1 visualizes this information, excluding examples with less than 2 occurrences.
For the research methods, half of the included articles chose the qualitative method, and the rest shared quantitative and mixed methods. Among the 102 articles, 57 articles used qualitative methods (accounting for 55.9%), 22 articles used quantitative methods (21.6%), and 23 articles used mixed methods (22.5%) (shown in Figure 2).

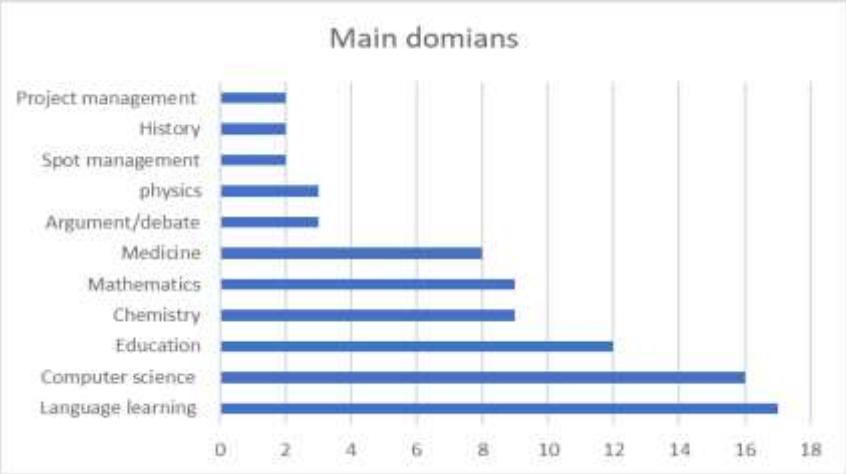

Figure 1. Main domains distribution.

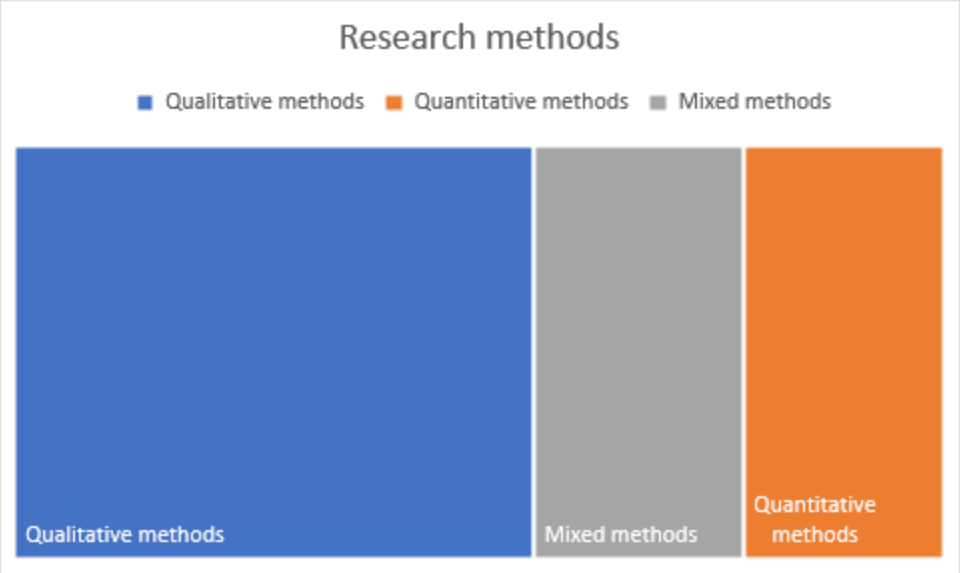

Figure 2. Research methods.

ChatGPT affordances in education

To better understand the construction of prompts in educational scenarios and the tasks at a general level, it is essential to extract the usage scenarios and affordances. More details can be found in Table 1. As shown in Table 1, ChatGPT provides more diverse affordances in two course phases: during the courses and after the courses.

Table 1 ChatGPT affordance in three teaching and learning stages

Stages (n)	ChatGPT support in teaching and learning (n)
Before the courses (32)	1) problem-solving/project-based learning support (4) 2) complex concept explanation or task expression simplification support (2) 3) plan/schedule generation/optimise support (10) 4) virtual role/specific sensitive case creation support (4) 5) test/questions generation/reframing support (8)

	6) complementary teaching materials support (4)
During the courses (39)	1) critical thinking support (3) 2) problem-solving/project-based/inquiry-based learning support (12) 3) reasoning support (1) 4) essay/report writing support (7) 5) dialogue/conversation/counselling skills support (4) 6) personalized tutor/coach support (6) 7) concept understanding (1) 8) self-reflection by writing support (2) 9) wrong solution explanation (1) 10) deeply thinking solution support (1) 11) code-tracing questions support (1)
After the courses (31)	1) essay feedback generation support (11) 2) design product feedback generation support (1) 3) code solution generation (1) 4) teaching materials skills tag (1) 5) report generate support (1) 6) code solution feedback generation support (4) 7) assessment feedback generation support (3) 8) peer comments feedback generation support (2) 9) rationales explanation for student answers (1) 10) students' responses/behaviour evaluation support (4) 11) students help seeking behaviour support (1) 12) coach task feedback support (1)

Data source for prompting and data collection for analysis

To gain experience in the data collection from the included articles, we summarized the commonly used data sources embedded into prompting (see Figure 3), and data collection methods (see Figure 4). When using and analysing ChatGPT, data collected from or related to course level was the main source for embedding context or task information into prompting, and data generated by ChatGPT was the main data source for further analysis of the ChatGPT outputs based on some prompts (rather than users' perceptions and performance impacted by ChatGPT). We found that the data sources mainly included course-level materials (46, 45.1%), ChatGPT output (22, 21.6%), university-level materials (12, 11.8%), lab collected materials (6, 6%), open access data (6, 6%), textbooks or domain knowledge or questions (4), national level materials (3), MOOCs or online platform (3). It indicates that data collected from or related to course level or data generated by ChatGPT are the main data sources embedded into prompting and analyzing the ChatGPT outputs. It was found that most of the articles relied on humans as evaluator to manually assess the ChatGPT outputs. The most common data collection and analysis path involved collecting conversation data between humans and ChatGPT and then comparing the differences between human-generated answers and ChatGPT-generated answers qualitatively or quantitatively.

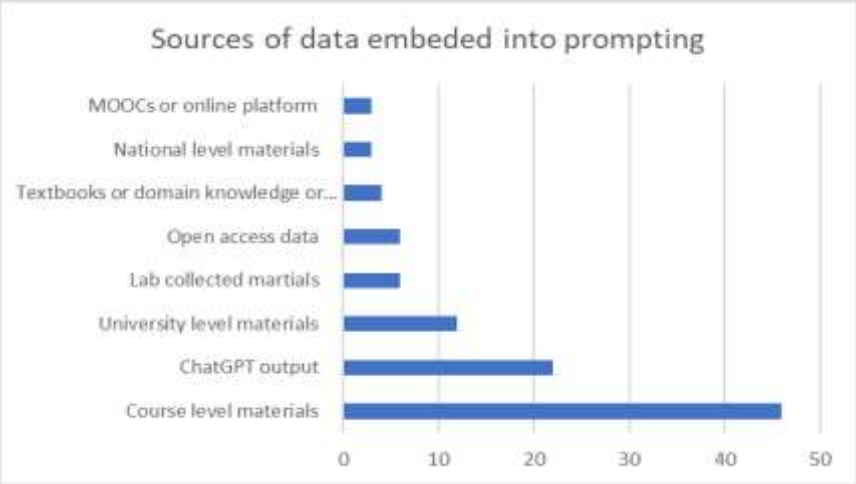

Figure. 3 Sources of data embedded into prompting.

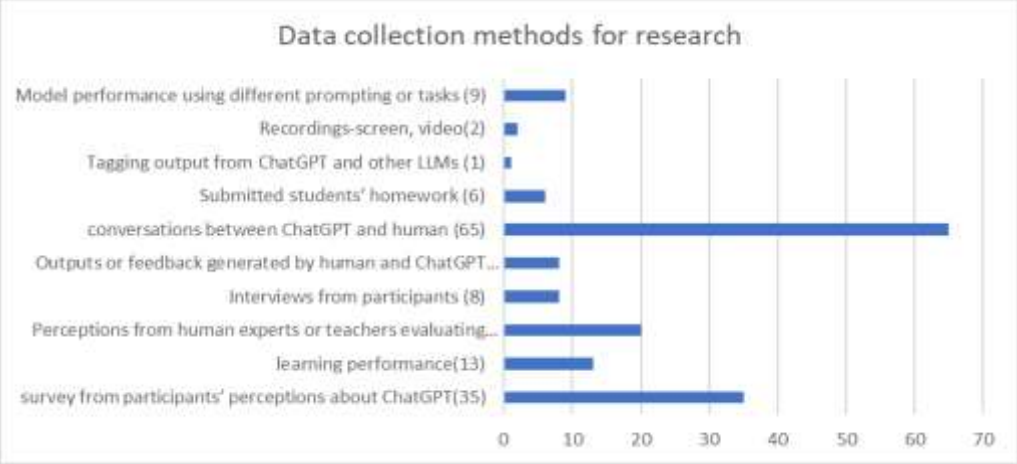

Figure 4. Common data collection methods from the included articles.
Note: Some articles employed more than one method.

4.2 RQ2

For educators

For educators, innovative teaching methods that connect the classroom with society using emerging technology can stimulate brainstorming for lesson activities, quickly collect complementary teaching and learning materials, test new instructional strategies with virtual students, optimize instructional scaffolding in class, generate parallel tests or exercises, formulate formative and final feedback, produce diverse positive comments for specific students, and summarize teaching workflow and progress. All of these would reduce repeated workload by using ChatGPT as an assistant.

Regarding pedagogical aspects, to enhance students' critical or reasoning thinking, teachers and stakeholders should consider investigating pedagogical approaches that leverage ChatGPT and similar GenAI tools in classrooms [4]. For example, integrating ChatGPT into learning by discovering, using ChatGPT to stimulate students deeply think about their solutions, combining other AI tools to generate multimodal products [6, 26]. Moreover, teachers play an important role in refining outputs from ChatGPT and strategically introduce ChatGPT and similar tools. With expertise in specific domains, teachers can use their prior knowledge and critical thinking skills to adapt ChatGPT outputs to deliver a more personalized educational context for students [1, 19].

Combining the use of GenAI tools like ChatGPT with different learning theories might produce different effects in teaching and learning [27]. This could benefit beginners by helping them quickly familiarize themselves with a field and experts' work efficiency. For example, information discovery tasks in a primary history course [26]. More theory examples can be found in a review of chatbot-assisted learning by [27]. More information about learner characteristics, instruction or design goal, content presentation strategies, assessment, primary concerns, pedagogical strategies, and application learning context in different theories can be found in the study by [28].

For students

For students, ChatGPT can provide around-the-clock personalized learning support and problem-solving guidance. For learning or teaching materials that are not customized for a specific domain, learners can use ChatGPT to reformulate the same question statements from a general domain to a specific domain, such as in a probability and statistics course, which could save time for this task [29], whether done by teachers or students.

Collaborating with ChatGPT or similar GenAIs, students need and would possibly improve digital literacy, AI literacy, and self-reflection skills [1, 19]. Users need to be aware of what tasks to complete or skills to develop, outline a ChatGPT usage workflow, and have alternative solutions when ChatGPT does not meet the task requirements, such as the AI-assisted workflow developed by [30]. Learners might get lost in the large number of generated outputs. The lengthy outputs generated from ChatGPT could overwhelm and stress learners with low prior knowledge [31]. In this situation, learners should be aware of the problem and adjust their prompts, such as limiting the output length and interpreting the outputs for a beginner with no prior knowledge. After iterations of collaboration with ChatGPT and self-summarizing, learners' digital literacy, AI literacy, and self-reflection skills would gradually improve.

5. Conclusion and Limitations

From the results of the source of data embedded in the prompting and data collection method, it can be found that diverse sources of data can be embedded in the prompt, and multiple methods can be used to analyse ChatGPT outputs. Course-level materials and ChatGPT outputs were commonly used. Moreover, users can develop and cumulate prompting archives by collecting prompts and discussing them with other users. This article found that data from conversations between ChatGPT and humans and surveys of users' perceptions were also methods to analyse the effectiveness of prompt construction.

As shown in ChatGPT affordances in education, GenAI tools take an assistant role to help learners complete tasks, but the important points are still on the cultivation of high-level knowledge and skills. It is deduced that the scaffolding of cognition and metacognition during the ChatGPT usage process would improve students' efficiency and effectiveness of prompting construction. This can be supported by the ChatGPT affordances in the education section. It was found that ChatGPT was adopted to provide support for critical thinking, problem-solving, self-reflection, and deep thinking.

Some limitations of this review can be found below. Firstly, the articles were searched from two big databases (WoS and Scopus), which is still possible to unintentionally miss some prompting-related empirical articles in the field of ChatGPT in education. To collect as much information as possible about the prompt details, the empirical studies in this review included not only quantitative articles but also qualitative articles. These qualitative articles might impact the generality of the regularity of the prompting and the output generated by using these prompts.

REFERENCES

[1] Mai D. T. T., Da C. V., Hanh N. V., "The use of ChatGPT in teaching and learning: a systematic review through SWOT analysis approach", Frontiers in Education, 2024, 9, 1328769.

[2] Kooli C., "Chatbots in education and research: A critical examination of ethical implications and solutions", Sustainability, 2023, 15(7), 5614.

[3] Ali J. K. M., Shamsan M. A. A., Hezam T. A., Mohammed A. A., "Impact of ChatGPT on learning motivation: teachers and students' voices", Journal of English Studies in Arabia Felix, 2023, 2(1), 41-49.

[4] Lo C. K., Hew K. F., Jong M. S. Y., "The influence of ChatGPT on student engagement: A systematic review and future research agenda", Computers & Education, 2024, 105100.

[5] Zhang P., Tur G., "A systematic review of ChatGPT use in K - 12 education", European Journal of Education, 2024, 59(2), e12599.

[6] Cronjé J., "Exploring the Role of ChatGPT as a Peer Coach for Developing Research Proposals: Feedback Quality, Prompts, and Student Reflection", Electronic Journal of e-Learning, 2023, 22(2), 1-15.

[7] Han J., Yoo H., Kim Y., Myung J., Kim M., Lim H., ... Oh A., "RECIPE: How to integrate ChatGPT into EFL writing education", Proceedings of the Tenth ACM Conference on Learning@ Scale, 2023, 416-420.

[8] Balse R., Kumar V., Prasad P., Warriem J. M., “Evaluating the Quality of LLM-Generated Explanations for Logical Errors in CS1 Student Programs”, Proceedings of the 16th Annual ACM India Compute Conference, 2023, 49-54.
[9] Cowan B., Watanobe Y., Shirafuji A., “Enhancing Programming Learning with LLMs: Prompt Engineering and Flipped Interaction”, Proceedings of the 2023 4th Asia Service Sciences and Software Engineering Conference, 2023, 10-16.
[10] Zheng Y. D., Stewart N., “Improving EFL students’ cultural awareness: Reframing moral dilemmatic stories with ChatGPT”, Computers and Education: Artificial Intelligence, 2024, 6, 100223.
[11] Meyer J., Jansen T., Schiller R., Liebenow L. W., Steinbach M., Horbach A., Fleckenstein J., “Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students’ text revision, motivation, and positive emotions”, Computers and Education: Artificial Intelligence, 2024, 6, 100199.
[12] Cohn C., Hutchins N., Le T., Biswas G., “A chain-of-thought prompting approach with llms for evaluating students’ formative assessment responses in science”, Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(21), 23182-23190.
[13] Hutt S., DePiro A., Wang J., Rhodes S., Baker R. S., Hieb G., ... Mills C., “Feedback on feedback: comparing classic natural language processing and generative AI to evaluate peer feedback”, Proceedings of the 14th Learning Analytics and Knowledge Conference, 2024, 55-65.
[14] Holderried F., Stegemann–Philipps C., Herschbach L., Moldt J. A., Nevins A., Griewatz J., ... Mahling M., “A Generative Pretrained Transformer (GPT)–Powered Chatbot as a Simulated Patient to Practice History Taking: Prospective, Mixed Methods Study”, JMIR medical education, 2024, 10(1), e53961.
[15] Lee G. G., Latif E., Wu X., Liu N., Zhai X., “Applying large language models and chain-of-thought for automatic scoring”, Computers and Education: Artificial Intelligence, 2024, 6, 100213.
[16] İpek Z. H., Gözüm A. I. C., Papadakis S., Kallogiannakis M., “Educational Applications of the ChatGPT AI System: A Systematic Review Research”, Educational Process: International Journal, 2023, 12(3), 26-55.
[17] Kumar S., Rao P., Singhania S., Verma S., Kheterpal M., “Will artificial intelligence drive the advancements in higher education? A tri-phased exploration”, Technological Forecasting and Social Change, 2024, 201, 123258.
[18] Crompton H., Burke D., “The Educational Affordances and Challenges of ChatGPT: State of the Field”, TechTrends, 2024, 68, 380–392.
[19] Ansari A. N., Ahmad S., Bhutta S. M., “Mapping the global evidence around the use of ChatGPT in higher education: A systematic scoping review”, Education and Information Technologies, 2023, 1-41.
[20] Liu P., Yuan W., Fu J., Jiang Z., Hayashi H., Neubig G., “Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing”, ACM Computing Surveys, 2023, 55(9), 1-35.
[21] Zamfirescu-Pereira J. D., Wong R. Y., Hartmann B., Yang Q., “Why Johnny can’t prompt: how non-AI experts try (and fail) to design LLM prompts”, Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, 1-21.
[22] Oppenlaender J., Linder R., Silvennoinen J., “Prompting ai art: An investigation into the creative skill of prompt engineering”, arXiv preprint arXiv:2303.13534, 2023.
[23] Dang H., Mecke L., Lehmann F., Goller S., Buschek D., “How to prompt? Opportunities and challenges of zero-and few-shot learning for human-AI interaction in creative applications of generative models”, 2022.
[24] Page M. J., McKenzie J. E., Bossuyt P. M., Boutron I., Hoffmann T. C., Mulrow C. D., ... Moher D., “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”, International journal of surgery, 2021, 88, 105906.
[25] Braun V., Clarke V., “Using thematic analysis in psychology”, Qualitative Research in Psychology, 2006, 3(2), 77–101.
[26] Murgia E., Abbasiantaeb Z., Aliannejadi M., Huibers T., Landoni M., Pera M. S., “ChatGPT in the classroom: A preliminary exploration on the feasibility of adapting ChatGPT to support children’s information discovery”, Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, 2023, 22-27.
[27] Zhang R., Zou D., Cheng G., “A review of chatbot-assisted learning: pedagogical approaches, implementations, factors leading to effectiveness, theories, and future directions”, Interactive Learning Environments, 2023, 1-29.

[28] Sun W., Pinkwart N., Li T., “Current state of learning analytics: a synthesis review based on the combination of activity theory and pedagogy”, Advancing the Power of Learning Analytics and Big Data in Education, 2021, 1-28.
[29] Einarsson H., Lund S. H., Jónsdóttir A. H., “Application of ChatGPT for automated problem reframing across academic domains”, Computers and Education: Artificial Intelligence, 2024, 6, 100194.
[30] Pham T., Nguyen T. B., Ha S., Ngoc N. T. N., “Digital transformation in engineering education: Exploring the potential of AI-assisted learning”, Australasian Journal of Educational Technology, 2023, 39(5), 1-19.
[31] Guo K., Wang D., “To resist it or to embrace it? Examining ChatGPT’s potential to support teacher feedback in EFL writing”, Education and Information Technologies, 2024, 29(7), 8435-8463.