Enhancing Learning with Student Teacher Augmented Response (STAR): A Multi-Agent Collaborative System for Nontraditional Education Settings

STAR: Enhancing Learning in Nontraditional Education Settings

Leveraging LLM-based Multi-Agent Systems in Education Beyond Conventional Classroom Settings

Xingyang He, C. He

YK Pao School, chrishe123321@gmail.com

ABSTRACT

With recent advancements in technology, integrating educational technologies into classrooms and schools has received considerable attention from both the industry and academia. Using technology to support nontraditional education settings (i.e., any form of learning outside the formal classroom setting) also should not be neglected, since it can enhance education's accessibility and help alleviate educational inequality. In this paper, we explored the limitations of current technologies (particularly ChatGPT) in facilitating nontraditional learning processes, which could be summarized as the content's lack of depth, time inefficiency, short memory, learners' impatience, lack of self-understanding, and inability to clearly express demand. To address these limitations, we developed a multi-agent collaborative system STAR (Student Teacher Augmented Response), proposing a novel multi-agent interaction framework, Student-Teacher Enactment, to align the generated response with the student's demand and better support student learning. Our findings show that STAR helps students gain a deeper understanding of the content learned, improves learning efficiency by reducing the time spent and effort required, and alleviates demotivating factors of learning through technology such as repetitive prompting.

CCS CONCEPTS • Multi-Agent Systems • Intelligent Tutoring System • Empirical Studies in HCI

1 INTRODUCTION

In recent years, the advancement of technology in day-to-day life, and the growing demand for online learning due to the global COVID-19 pandemic, have spurred the development of various educational technologies [25]. Such technologies mainly focus on supporting learning in classrooms, a traditional education setting [14]. These include supporting student learning via Intelligent Tutoring Systems and learner modeling [1], supporting teaching with data or machine learning [40], enhancing learning accessibility via massive open online courses (MOOCs, such as Khan Academy [70] and Coursera [71]).

Despite the growing body of literature on technological applications in traditional classroom settings, there is a notable gap in research regarding technologies in nontraditional education settings (i.e., learning outside formal education) [14]. Exploring technology's potential in nontraditional, outside-classroom learning is essential [27], since it can enhance education accessibility for adult learners [55] and alleviates the issue of educational inequality by providing learning opportunities for underprivileged students [6].

The advent of ChatGPT [72], a large language model developed by OpenAI, has stimulated a new round of evolution in the field of educational technologies. Due to its conversational abilities, ChatGPT has been and is continuously being integrated into traditional classrooms to enhance student learning in a variety of domains including language learning [4], knowledge acquisition [12, 13], providing model answers for higher education exam questions [14, 15], etc. However,

ChatGPT's applications in nontraditional settings are still quite limited, mainly focusing on integrating it with online learning platforms and MOOCs, such as Khan Academy's AI teacher tool, Khanmigo [73].

Recognizing these gaps and the potential for broader applications, this paper aims to leverage the power of ChatGPT, multi-agent collaboration, and prompt engineering to support learning, particularly self-directed and outside-classroom learning, in nontraditional settings. In this paper, we aim to answer the following research questions:

RQ1: What are the limitations of current technologies in supporting student learning in nontraditional education settings?

RQ2: How can we design and develop a chatbot (STAR) that effectively supports students' learning in nontraditional education settings?

RQ3: What are the preliminary effects of STAR on students' learning in nontraditional education settings?

2 BACKGROUND

2.1 Technology in Traditional Education

2.1.1 Importance of Technology in Education

Technology has transformed the classroom from a passive learning space to an interactive learning environment [61]. The learning process becomes more engaging and interactive with the aid of technologies, including gamification tactics [31, 53], 3D simulations [47], online collaborative platforms [30], etc. Education resources have become more accessible through the establishment of online learning platforms and MOOCs [35]. The management and efficiency of the educational system also improved drastically with the emergence of learning management systems [22, 28, 58].

2.1.2 Current Technology in Traditional Education Settings

Various technologies have been and continue to be developed by industry and academia to support learning in traditional education settings (i.e., within the classroom environment). Technologies such as augmented reality (AR) [2, 20, 47], artificial intelligence (AI) [36, 62], intelligent tutoring systems [5, 11, 50], etc., have been studied thoroughly and have shown success in enhancing students' motivation and learning [37, 54]. To list a few examples, audience response systems have been developed to enhance classroom interactivity by allowing learners to provide instantaneous responses to teachers' questions [45, 51]. Additionally, computer vision techniques have been used in cooperation with the Internet of Things to provide teachers with data and real-time analysis of students' posture, attention, emotion, etc., aiding the delivery of classes and lectures [56, 74].

2.2 Nontraditional Education Settings

In addition to traditional education settings, educational technology's application in nontraditional settings has received increasing attention over the last decade.

2.2.1 Definition of Nontraditional Education Settings

Nontraditional education settings can be defined as any environment of learning outside formal education (i.e., K12 and college education) [13]. It can be characterized by its flexible teaching content (in contrast to K12 formal education, which needs to follow syllabi published or approved by the Department of Education [42]) and its independence (i.e., it can happen anywhere at any time, without a human teacher). According to this definition, nontraditional education settings

encompass a wide range of possibilities and technologies, including, but not limited to, company training, self-directed learning using YouTube videos, reading books at a library, and touring a museum [9].

2.2.2 Importance of Nontraditional Education Settings

Nontraditional education significantly enhances nontraditional learners' (i.e., adult learners, part-time students, etc.) access to education and educational resources [55]. It alleviates the issue of education inequality (i.e., the disparities in access to quality education based on socioeconomic status [62]) by granting underprivileged learners learning opportunities and additional resources [6]. For adult learners, the flexibility of nontraditional education settings which allow learners to control their own pace and schedule has the additional benefit of balancing education with their work and family responsibilities [52]. Furthermore, the flexible content delivered in this setting ensures that the education received is relevant to the learners' career or personal growth, enhancing learners' intrinsic motivation and interest in learning [48].

2.2.3 Current Technology in Nontraditional Education Settings

With the increasing focus on equality and accessibility around the globe, there is a growing amount of research on supporting the different nontraditional education settings [15]. Due to the pandemic and the popularization of online learning, recent efforts have been focused on online tutoring systems and learning platforms [44]. Utilizing AI technologies, researchers have focused on modeling learners' characteristics, strengths, and weaknesses [4, 29], predicting learners' performances and progress [34, 64], and recommending specialized courses to personalize students' learning paths [5, 30].

2.3 ChatGPT and Education

2.3.1 ChatGPT

ChatGPT is an Artificial Intelligence (AI) chatbot developed by OpenAI based on the transformer architecture [63]. Given an input text, it utilizes regression techniques to repeatedly predict the subsequent token, producing responses that understand the input and highly emulate that provided by a human [59]. Since its first release in November 2022, it has received considerable attention from universities worldwide, with professors and administrators actively exploring its potential to transform traditional pedagogical and educational approaches [69].

2.3.2 ChatGPT in Traditional Education Settings

Currently, ChatGPT's applications in classrooms focus on two types of users, teachers and students, aiding the delivery of subject-specific knowledge. For teachers, emphasis is put on helping develop course materials [12, 19] and auto-grading student's responses [7, 57]. For students, its use can be further divided into two categories: language and non-language learning. In language learning, ChatGPT can act as a conversational partner, improving students' ability to interact in the language environment they are learning [4, 35, 66]. Additionally, it can also provide evaluative feedback on students' writing [60] and generate ideas when composing writings [67]. In subjects other than language acquisition, ChatGPT is utilized to help with students' unanswered questions in class [12, 13, 60] and provide sample responses for exam questions [14, 15].

2.3.3 ChatGPT in Nontraditional Education Settings

Outside classrooms (e.g., in-home self-study), efforts are put into integrating LLM-based agents with online learning platforms, such as Khan Academy's AI-powered tutor Khanmigo [73], which integrates Khan Academy's 10,000-plus educational content and provide step-by-step guidance for students [75]. Efforts have also been made to develop

personalized, subject-specific AI tutors using fine-tuning and prompt engineering techniques [62, 63, 64]. An example of this is Iris [3], an AI tutor who leverages chain-of-thought processing, post-generation self-check, and external code databases to support students' independent knowledge acquisition in the field of computer science.

3 METHODOLOGY

3.1 Study 1: Exploring Uses and Limitations of Current Technologies in Nontraditional Education Settings

3.1.1 Data Collection – Interview and Contextual Inquiry

To explore ChatGPT's usage and potential application in nontraditional settings, we first aimed to understand (1) learners' current approaches toward nontraditional learning and (2) learners' perception of ChatGPT as a learning tool. To achieve objectives (1) and (2), we conducted three sets of semi-structured interviews and contextual inquiries following the high-level protocol included in the Appendix. Participants were recruited from high schools and universities in Shanghai and Boston. The three participants are male (2) and female (1) high school or undergraduate students ranging from 18 to 20 years old. All participants were pursuing some form of nontraditional learning (e.g., self-directed learning or online learning) and had diverse interests. Two of the participants regularly used ChatGPT in their learning process, while the third participant had experience using other nontraditional education technologies, such as the MOOC platform Coursera.

The study was structured as follows. First, we introduced our research objective and collected contextual information on participants' existing nontraditional learning practices. Then, we performed a contextual inquiry where we gave each participant a research task that resembled their daily outside-classroom learning activities (e.g., learning about the impacts of Japan's ultra-loose monetary policy on societal well-being). We asked the participants to think aloud (i.e., speak aloud any words in their mind as they complete a task [8]) and observed their approach to the task. The contextual inquiry session lasted 20 minutes. We restricted participants' use of ChatGPT for the first 10 minutes and then lifted this restriction. During the session, we occasionally asked questions to the participant, aiming to understand the motivation and rationale behind their actions [23]. After the session, we asked participants to comment on and evaluate ChatGPT's performance in the session and their daily use. We recorded the interviews and contextual inquiries and transcribed the data using Podium [76]. In total, we collected 2.25 hours of recorded data.

3.1.2 Data Analysis – Interpretation Session and Affinity Diagramming

We used two standard techniques from Contextual Design to analyze the qualitative data collected: Interpretation Sessions and Affinity Diagramming [24]. During the Interpretation Session, we followed the conventional approach [43], worked with my faculty advisor (an assistant teaching professor), and documented our observations, ideas, and insights yielded from the interviews onto sticky notes, known as our first-level insights. During this process, we acquired a multi-faceted understanding of the data collected [65]. Then, we leveraged Affinity Diagramming to cluster the 165 first-level insights into 38 second-level insights based on the similarity of the observations' contents. We performed the clustering on the online platform Miro [77] and, following similar approaches as before, grouped the 38 second-level insights into 6 third-level insights. These third-level insights represent our understanding of the technological and human factors that hinder learning in nontraditional education settings, and we presented them in section 4.1.

3.2 Chatbot Design – STAR

Based on the insights yielded from the Affinity Diagramming, we designed a multi-agent collaborative system STAR (Student-Teacher Augmented Response, Figure 1) that aims to facilitate learners' self-directed learning outside the classroom by acting as a virtual tutor that delivers knowledge in a specific and personalized manner.
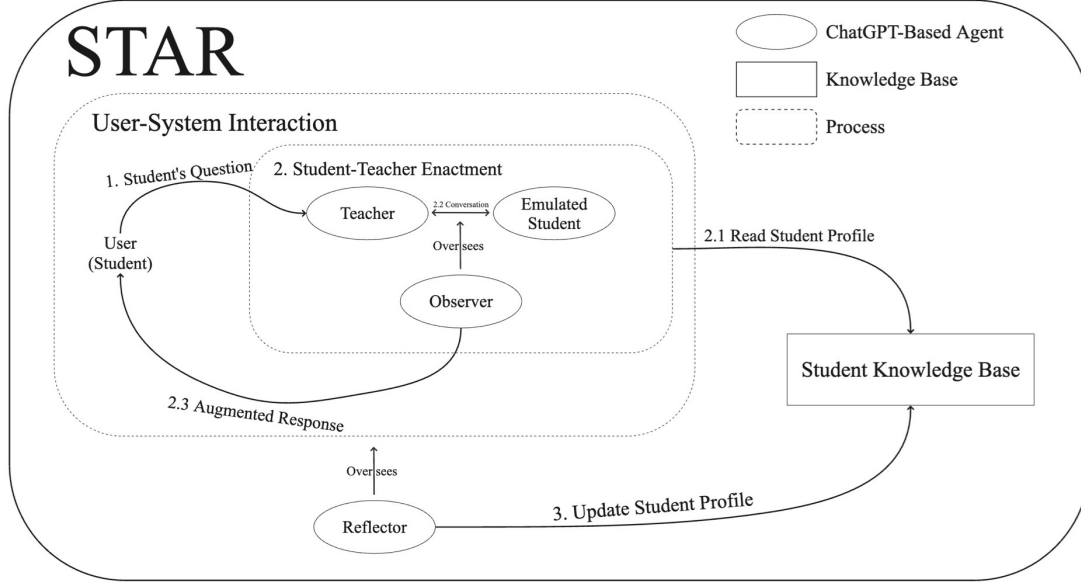

Figure 1: STAR System Diagram

STAR consists of three basic entities: User-System Interaction, Reflector, and Student Knowledge Base. User-System Interaction is further divided into the user and the Student-Teacher Enactment process. All prompts used in STAR as system instructions for different LLM-based agents are included in the Appendix.

The system's workflow can be described as follows. First, the human user (a learner) asks the chatbot a question. This question serves as the input to the Teacher agent. A student profile documenting the user's learning preferences (e.g., prefers concise responses), conversational style (e.g., often requests specific examples), and other relevant information, is obtained from the Student Knowledge Base and given to the Teacher agent as another input (the construction of student profiles is discussed latter). With this information, the Teacher agent provides an answer to the user's question tailored to their needs, which is called the *initial response*. This response, as well as the user's student profile, is then inputted to an Emulated Student agent. The Emulated Student agent can be seen as a duplication of the user. Its role is to enact the human user (i.e., act and converse in the same way as the human user) by using the data in the user's student profile. The Emulated Student agent engages in a conversation with the Teacher agent, and the Observer agent oversees this conversation. It uses the follow-up questions raised by the Emulated Student agent to identify unclear points or weaknesses in the Teacher agent's initial response. The Observer agent then improves the initial response accordingly and outputs the augmented response to the human user. Typically, the human user will respond with a follow-up question. The process outlined above, where the user interacts with the Student-Teacher Enactment process, is denoted as User-System Interaction.

The second part of the STAR architecture is the Reflector. A Reflector agent oversees User-System Interaction. Whenever the user responds to Student-Teacher Enactment's output (i.e., the augmented response), the user's response is inputted into the Reflector agent. The Reflector agent is also given the user's student profile and the conversation history between the Teacher and Emulated Student agent during Student-Teacher Enactment. The Reflector agent then compares and contrasts the user's response with the Emulated Student agent's response, observes and extracts high-level insights into the user's characteristics, preferences, personality, etc., and updates the user's student profile in the Student Knowledge Base. The next time the user asks a question, Student-Teacher Enactment will retrieve the updated student profile from the Student Knowledge Base.

To implement the STAR system, we utilized Microsoft's AutoGen, a framework that facilitates the development of multi-agent workflows [78]. The system was developed in Python, and the agents are based on OpenAI's LLM gpt-4o-2024-05-13 [79].

3.3 Study 2: Exploring STAR's Preliminary Effects on Learners

To explore whether STAR addresses some of the limitations of existing technologies, which we learned from the Affinity Diagramming in our first study, and helps support learning in nontraditional education settings, we designed the following experiment. The experiment consists of three stages: pre-session knowledge test, technology-enhanced learning session, and post-session knowledge test. First, participants are given a learning focus and learning objectives, such as the philosophical concept of Lacan's "big Other" [29]. Each participant takes a pre-session knowledge test on that topic, which includes multiple-choice, short-answer, application, and critical-thinking questions. Then, participants are randomly assigned to one of two experiment groups, learning with ChatGPT and learning with STAR, and begin learning their assigned topic. They end their learning session when they feel satisfied with their learning results. During the learning session, data such as the total number of questions raised by the participant, the total number of words inputted, the amount of time spent, etc., are recorded. After the learning session, each participant takes a post-session knowledge test that has a similar structure and content to the pre-session test. Then, we presented the internal conversation history between the agents to the participants and engaged in a small discussion with them. We included an example of the pre- and post-session knowledge tests we used, their mark schemes, and the detailed experiment protocol in the Appendix.

To ensure the application scope of the experiment's results, we selected six learning focuses that incorporate a wide range of subject areas, such as philosophy, economics, psychology, etc. We recruited 4 participants (2 male and 2 female, ranging from 18-21 years old) following the same process outlined in section 3.1.1, and we randomly assigned three learning focuses to each participant. In total, 12 learning sessions (6 using ChatGPT 4o and 6 using STAR) with their corresponding knowledge tests are performed, recorded, and analyzed. To avoid the cold-start problem (i.e., not having enough information and data on new users [41], which affects STAR's performance as we cannot construct detailed and accurate student profiles), we asked the participants to provide their conversation history with ChatGPT. Then, based on the conversation history, we prompted the Reflector agent to construct a student profile for each participant that follows the format of student profiles we designed.

4   RESULTS

4.1   Findings from Study 1

This section outlines the 6 third-level insights yielded from Affinity Diagramming. They are the limitations of current technology (particularly ChatGPT) in supporting student learning in nontraditional education settings.

4.1.1 Technological Factors: Lack of Depth, Time Inefficient, Short Memory

First, in our study, we found that ChatGPT may arrive at a "depth bottleneck" where it cannot delve deeper into a specific topic. This is often characterized by multiple rounds of conversation where the user provides specific instructions for improvement but ChatGPT continues to output similar or repetitive answers. This limited depth hinders ChatGPT's knowledge delivery in subjects that rely on critical analysis and nuanced interpretation, such as Philosophy and Literature. This is demonstrated by several participants' actions during Contextual Inquiry. They switched to using Google and subject-specific websites such as the Stanford Encyclopedia of Philosophy [80] after a few interactions with ChatGPT that helped them gain a general overview of the subject area.

Second, we found that ChatGPT relies on multiple rounds of interactions with the user to align its output to the user's needs and demands. Often, ChatGPT's initial response to the user's question is generic and lacks detail. Then, by asking follow-up questions, the user gradually focuses ChatGPT's attention on what they want to learn about. This iterative approach suits the transformer model and the attention mechanism that forms the basis of ChatGPT. However, to the user, this approach consumes time since they need to wait for ChatGPT to generate its output and provide feedback accordingly. Multiple participants expressed that "the process of guiding ChatGPT to understand what you really want is long and treacherous" and "you don't feel like you are learning when you spend half of the time telling it [ChatGPT] what you want." This signifies how needing to interact multiple times with ChatGPT may demotivate users from learning.

Third, we found that ChatGPT has a short memory for users' instructions. For instance, in our study, a participant stressed "Use specific examples to support your statement". However, after three rounds of conversation, ChatGPT returned to a generic output style. This inconsistency means that users need to constantly prompt ChatGPT to make sure it doesn't forget its instructions. This need for repetitive prompting can be particularly frustrating, as users may feel that their initial efforts to guide the conversation are wasted. Participants in our study noted that "You have to keep reminding it [ChatGPT] of what you asked for, which is tiring and breaks the flow of learning." The repetitive prompting discourages users from relying on ChatGPT as an educational tool

4.1.2 Human Factors: Impatience, Lack of Self-Understanding, Unclear Expression

Fourth, human users' impatience when using ChatGPT also hinders its knowledge delivery. Often, users do not consider ChatGPT as a tutor that needs to be respected. For instance, we observed that participants often do not read ChatGPT's full response. Instead, they will raise a question the instant they see one. This strong sense of purpose and frequent user interaction breaks the coherence of gradual knowledge acquisition, hindering knowledge delivery by ChatGPT. Moreover, frequent interaction increases the time users spend waiting for ChatGPT to respond, inducing "boredom and a sense of meaningless" amid knowledge acquisition that demotivates users to continue learning.

Fifth, in nontraditional education settings that do not have human teachers, students lack the ability to identify gaps in their knowledge and understanding. In the classroom, assessments, tests, teacher's direct feedback, etc. help students understand their weaknesses, allowing them to practice and improve accordingly. However, without human teachers' support, students may not have the capability of self-reflection. We observed cases where the participant felt confident in their understanding, but questions we raised indicated that they didn't or misunderstood important concepts in that area. Existing technologies such as MOOC platforms attempt to incorporate personalized assessments into the courses to facilitate learners' self-understanding. However, as stated by participants, their effects aren't significant as it is hard to personalize assessments to an appropriate level of difficulty.

Sixth, as stated by several participants, learners may struggle to "articulate their inner thoughts", making it difficult for ChatGPT to understand their demands and produce responses that effectively address them. In traditional classroom settings, human teachers are able to understand and discern what students want or don't understand based on their experience. However, ChatGPT lacks the semantic understanding and intention-prediction skills needed to understand students' demands when they are not accurately expressed. In addition, ChatGPT's architecture means that it relies on the user's input to generate output. Thus, in nontraditional education settings, users' inability to articulate their demands may hinder technologies like ChatGPT's knowledge delivery.

4.2 Findings from Study 2

After reviewing our findings, we prioritized addressing the first, second, third, and sixth limitations. Our decision is based on the participants' feedback, who identified them as having the most direct and significant impact on their learning experience. Our proposed solution is a new multi-agent interaction framework, Student-Teacher Enactment, and a multi-agent collaborative system STAR in which we implemented the proposed framework. As for the fourth and fifth limitations, we decided to leave them for future research. These limitations are related to human factors that may not be easily addressed or solved with technology itself.

In Study 2, we tested STAR's preliminary effects with 4 volunteer participants and recorded the results in the following section.
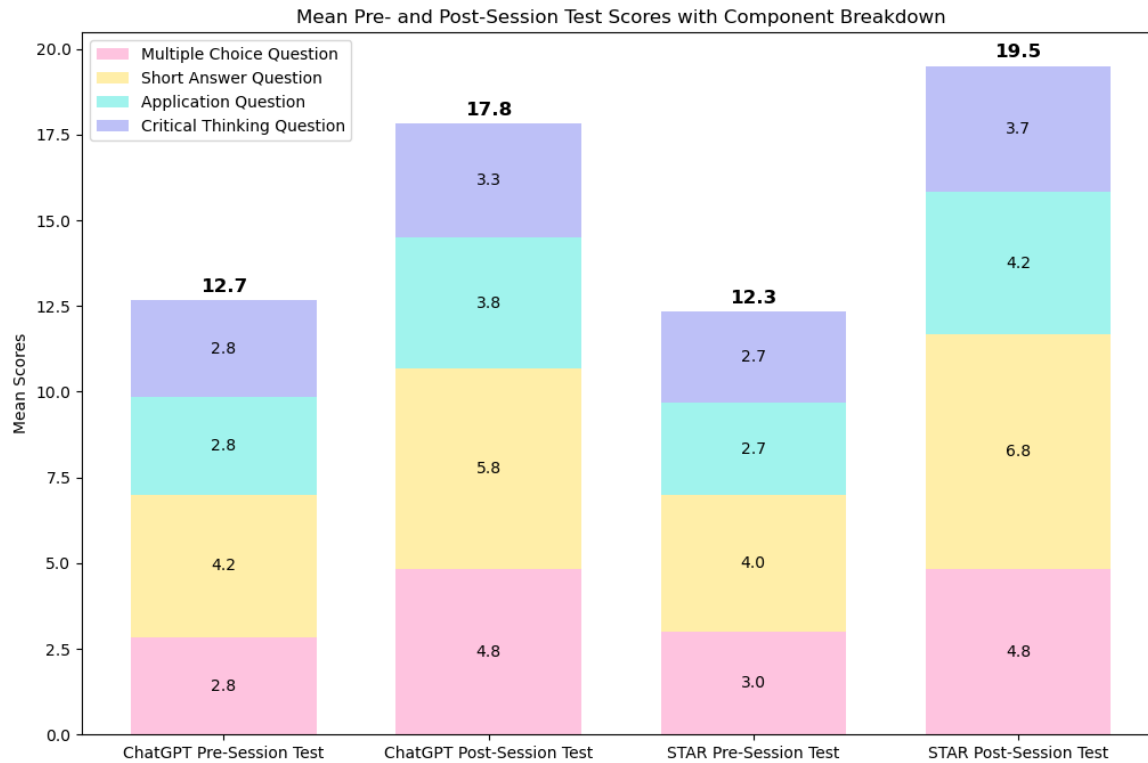
4.2.1 Effectiveness in Supporting Learning

Figure 2: Comparison of pre-session and post-session test scores between learning sessions using ChatGPT and STAR

Based on our findings, STAR shows competence in supporting learning in nontraditional education settings. In this study, we used the scores of pre- and post-session knowledge tests as the metric for evaluating STAR's performance. As shown in Figure 2, participants score similarly in the pre-session knowledge tests, with the ChatGPT and STAR groups scoring 12.7 and 12.3 out of 22 on average, respectively. After the learning session, participants who used STAR showed a higher increase in score, a percentage increase of 58.5%, than those using ChatGPT, which had a percentage increase of 40.2%. This illustrates and proves STAR's ability to deliver knowledge and support learning in nontraditional education settings.

4.2.2 Enhanced Depth of Understanding

Our findings also showed that STAR helped participants gain a deeper understanding of the learning topic than their peers that used ChatGPT. This is shown in the change in scores of critical thinking questions (CTQ). They are a specific type of questions that aim at assessing students' understanding depth of a topic by challenging them from the meta-level and prompting reflections on the underlying assumptions, cognitive biases, etc. of the topic. The average CTQ score for each test is shown in the bars' purple section in Figure 2. Figure 3 shows the percentage change in CTQ score between pre- and post-session tests for learning sessions using ChatGPT and STAR.
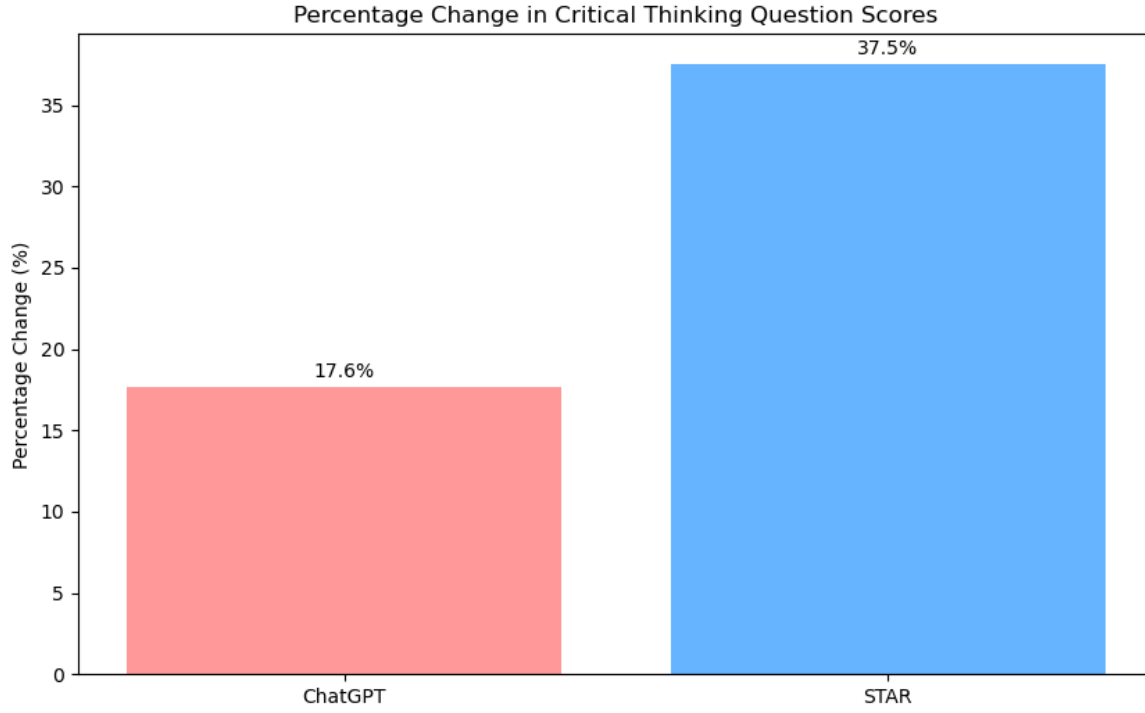
Figure 3: Comparison of percentage change in critical thinking question's score between learning sessions using ChatGPT and STAR

As we can see, participants using STAR in learning sessions have a higher increase in CTQ score, 37.5%, compared to participants using ChatGPT. This indicates that STAR effectively helps learners gain a deeper understanding of the topic. We believe this is mainly due to the Student-Teacher Enactment framework in STAR. Specifically, we and the participants observed multiple times where the Emulated Student agent raised "unexpected but thought-provoking" questions to the Teacher agent. When the Observer agent tried to address them in the augmented response, the produced learning content's depth increased. The high-level thinking incorporated in these augmented responses further prompted participants' (learners') reflection and evaluation on the topic.

4.2.3 More Efficient Use of Time and Effort

We also found that STAR significantly reduced the time and effort needed for learners to gain a satisfactory understanding of the topic. This is shown in measuring the following metrics:

$$Percentage\ Change\ in\ Score\ per\ Time\ Spent = \frac{\frac{Post\ Session\ Score - Pre\ Session\ Score}{Pre\ Session\ Score}}{Time\ Duration\ of\ Learning\ Session\ (Second)}$$

$$Percentage\ Change\ in\ Score\ per\ Word\ Inputted = \frac{\frac{Post\ Session\ Score - Pre\ Session\ Score}{Pre\ Session\ Score}}{Number\ of\ Words\ Inputted}$$

The two metrics measured the effect of a second spent and a word inputted (to ChatGPT or STAR) on the student's performance.
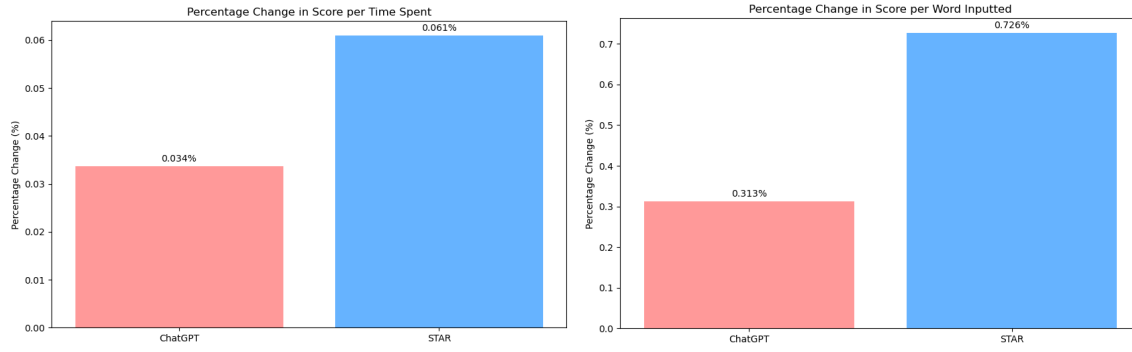
Figure 4: Comparison of percentage change in score per time spent and word inputted between learning sessions using ChatGPT and STAR

As we can see, STAR showed a better performance on both metrics. This means that compared to ChatGPT, learning with STAR can achieve a similar outcome with less time and effort. The better utilization of time may alleviate the sense of impatience felt by some users and enhance students' learning motivation. The less input required from users meant that learners could focus more attention on learning itself, enhancing the learning experience. In addition, multiple participants expressed that STAR is better at "understanding what [they] want" than ChatGPT. This aligns with our observation where participants using STAR raised a lower number of questions on average than participants using ChatGPT: 5.17 compared to ChatGPT's 8.83. This indicates that users need to spend less effort to steer STAR's response to what they want, which further enhances the learning experience and alleviates the demotivating factors.

We believe the positive outcomes mentioned above are the results of the Student-Teacher Enactment framework's ability to predict the user's demand. As mentioned, before STAR outputs a response, it first calls the Emulated Student agent to simulate and predict the user's potential reactions and the possible follow-up questions they may raise. The Observer agent addresses these questions in the augmented response. If the Emulated Student agent raises the same points as the participants (which is our aim), the need for users to further clarify their requests is eliminated, and the learning process' time and input efficiency are increased.

4.2.4 Less Repetitive Prompting

Based on observations and empirical evidence, we concluded that STAR requires less repetitive prompting by the user. This is most vividly shown in a participant's remark "I enjoy seeing the agent inputting the prompts that I usually need to input myself for me." Requiring less repetitive prompting alleviates users' boredom and "sense of meaningless" that may arise from their repeated input of an instruction.

We believe this is due to the prevalence of the user's student profile in STAR. It is included in the system prompt of every agent, constantly reminding the system of the user's preferences like "stay concise". Thus, the need for the students to manually input these instructions is reduced. Moreover, the user's learning experience is more coherent as they don't need to spend effort in inputting instructional prompts such as "use analogies" that aren't related to the learning content.

11

4.2.5 Decreased Need for Clear Expression

Our research also indicated that the Student-Teacher Enactment framework reduces the need for clear and explicit instructions from users. This might be attributed to the high expressiveness of student profiles and the advanced emulative capabilities of LLMs, which enable the Emulated Student agent to effectively assume the user's role. Upon reviewing the agent's internal conversation history, several participants noticed that the questions they had after reading the Teacher agent's initial response were already raised by the Emulated Student agent. In some instances, participants even observed that the Emulated Student agent clearly expressed a point they found confusing but couldn't articulate themselves. This indicates that, as the Student-Teacher Enactment framework can "predict" what the user is confused about, the need for the user to accurately express their demand is reduced, and the barriers for learners to use STAR as a learning tool are lowered. Thus, the accessibility of knowledge and tools for knowledge delivery are increased.

5   DISCUSSION AND FUTURE STEPS

While Large Language Models' (LLMs, e.g., ChatGPT) versatility led to their application in various domains, their use in education, especially supporting learning in nontraditional education settings, has received limited attention from academia and industry. The issues of lacking depth, time inefficiency, repetitive prompting, and learners' inability to clearly express their demands further hindered the potential of technology to provide additional educational resources for underprivileged students in nontraditional education settings. To address these issues and fill the gap, we proposed a new multi-agent interaction framework called Student-Teacher Enactment. This framework has been implemented in our multi-agent collaborative system, STAR, and we conducted testing to evaluate its effects. The results from Study 2 (our pilot study of STAR) showed that STAR allows users to gain a deeper understanding of the learning topic, reduces the time and effort required to obtain a satisfactory response, alleviates the issue of repetitive prompting, and decreases the need for users' clear expressions. By improving technology's ability to deliver personalized knowledge and addressing factors that demotivate learning, STAR helps make high-quality educational resources more accessible to students in nontraditional education environments, bridging the gap between traditional and nontraditional learners. More importantly, STAR facilitates the spread and democratization of knowledge through nontraditional education settings, preventing knowledge from being monopolized by a few privileged entities that are overly powerful in traditional education settings.

5.1   Limitations of STAR's Design

5.1.1 Cold-Start Problem

First, there is the cold-start problem, where it is hard to construct new users' student profiles. In Study 2, we performed analysis of participants' chat history with ChatGPT to construct their profiles. However, this approach raises privacy issues that make it unrealistic for real-life applications. In the future, we will explore methods to address this problem, such as hardcoding generic student profiles that can represent a group of users, or designing a pre-use test that helps us collect students' information and preferences.

5.1.2 Long Conversation History that is Distracting

Second, the long conversation history (between the Teacher and Emulated Student agent) given to the Observer agent may distract it from targeting the user's question in the augmented response. An example of this is shown in Figure 5, where the Observer agent includes details on the "blood diamond" phenomenon (because the Emulated Student agent asks a

follow-up question on it) while the user asks for a high-level overview of "diamonds' significance to developing areas of the globe".
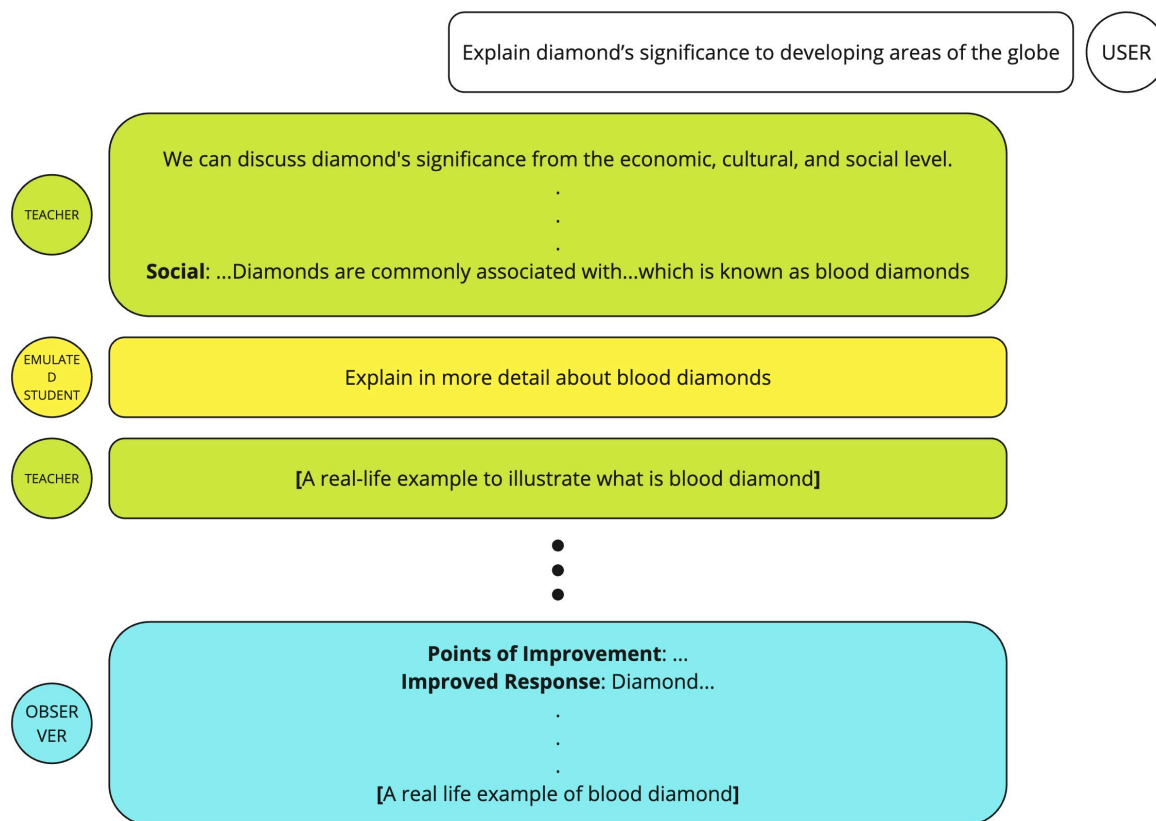

Figure 5: An example of internal conversation history between agents in Student-Teacher Enactment, illustrating how long conversations distract the Observer agent from the user's initial question

We observed that the possibility for this to happen has a positive correlation with the maximum round of conversation allowed between the Teacher and Emulated Student agent. In future works, we will run studies to fine-tune this hyperparameter and find the balance point between yielding maximum information and remaining focused on the user's initial question.

5.1.3 Performance-Cost Trade-Off

Third, STAR and its multi-agent interaction framework introduces a performance-cost trade-off, where responses that better address the user's demand requires more interaction between agents and higher token costs. These costs may not be negligible for nontraditional learners like part-time students, limiting STAR's usage and performance.

5.2 Limitations of Experiment Designs

5.2.1 Low Participant Diversity

First, our studies' participants are all high school or undergraduate students. This means that our findings may not be representative of the diverse learning needs present in nontraditional education settings. In the future, we will incorporate a more diverse participant group that allows the voices of other types of learners to be heard.

5.2.2 Choice of Metric

Second, the metrics we measured, such as the percentage change in score per word inputted, may be influenced by factors other than the system design, such as the participant's language style. These noises may skew the data collected, making it an inaccurate representation of STAR's performances. In future works, we will try to develop new metrics that more accurately measure quantities such as a system's time and input efficiency.

5.2.3 Limited Scope of Content

Third, the content covered in our study was limited to 6 subjects, which may not fully represent the wide range of material that learners typically engage with. Future research should include a broader array of subjects to determine if STAR's effectiveness is consistent across different areas of study.

5.3 Future Steps

Our future steps center around the unaddressed limitations identified in Affinity Diagramming. We will explore methods to help learners identify gaps in their knowledge, such as adding personalized assessments to the learning process. In addition, we will try to enhance STAR's capability by adding function-calling agents into the workflow, such as using Tavily's API [81] to browse the internet which can help the Teacher agent collect real-life examples.

Furthermore, we will perform an ablation study to explore the contribution of each entity to STAR's performance, which better guides our future design iterations. We will also perform benchmark testing of STAR with other AI tutors that aim to support self-directed learning, such as Tutor AI [82] and Khanmigo [75], to comprehensively evaluate STAR's performance and identify areas for further improvement.

REFERENCES

[1]   Abir Abyaa, Mohammed Khalidi Idrissi, and Samir Bennani. 2019. Learner modelling: systematic review of the literature from the last 5 years. *Educ. Technol. Res. Dev.* 67, 5 (October 2019), 1105–1143. https://doi.org/10.1007/s11423-018-09644-1

[2]   Murat Akçayır and Gökçe Akçayır. 2017. Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educ. Res. Rev.* 20, (February 2017), 1–11. https://doi.org/10.1016/j.edurev.2016.11.002

[3]   Patrick Bassner, Eduard Frankford, and Stephan Krusche. 2024. Iris: An AI-Driven Virtual Tutor For Computer Science Education. https://doi.org/10.1145/3649217.3653543

[4]   Jose Belda-Medina and José Ramón Calvo-Ferrer. 2022. Using Chatbots as AI Conversational Partners in Language Learning. *Appl. Sci.* 12, 17 (August 2022), 8427. https://doi.org/10.3390/app12178427

[5]   Elizabeth Emily Carter. An Intelligent Debugging Tutor For Novice Computer Science Students.

[6]   Ismail Celik, Muhterem Dindar, Hanni Muukkonen, and Sanna Järvelä. 2022. The Promises and Challenges of Artificial Intelligence for Teachers: a Systematic Review of Research. *TechTrends* 66, 4 (July 2022), 616–630. https://doi.org/10.1007/s11528-022-00715-y

[7]   Li-Hsin Chang and Filip Ginter. 2024. Automatic Short Answer Grading for Finnish with ChatGPT. *Proc. AAAI Conf. Artif. Intell.* 38, 21 (March 2024), 23173–23181. https://doi.org/10.1609/aaai.v38i21.30363

[8]   Elizabeth Charters. 2003. The Use of Think-aloud Methods in Qualitative Research An Introduction to Think-aloud Methods. *Brock Educ. J.* 12, 2 (July 2003). https://doi.org/10.26522/brocked.v12i2.38

[9]   Joseph C. Chen. 2017. Nontraditional adult learners: The neglected diversity in postsecondary education. *Sage Open* 7, 1 (2017), 2158244017697161.

[10]   Yu Chen, Scott Jensen, Leslie J. Albert, Sambhav Gupta, and Terri Lee. 2023. Artificial Intelligence (AI) Student Assistants in the Classroom: Designing Chatbots to Support Student Success. *Inf. Syst. Front.* 25, 1 (February 2023), 161–182. https://doi.org/10.1007/s10796-022-10291-4

[11]   Cristina Conati. Intelligent Tutoring Systems: New Challenges and Directions.

[12]   Grant Cooper. 2023. Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *J. Sci. Educ. Technol.* 32, 3 (June 2023), 444–452. https://doi.org/10.1007/s10956-023-10039-y

[13]   Patricia Cross. THE NEEDS OF NON-TRADITIONAL LEARNERS AND THE RESPONSES OF NON-TRADITIONAL PROGRAMS.

[14]   Adolph Delgado, Liane Wardlow, Katherine Mcknight, and Kimberly O'Malley. 2015. Educational Technology: A Review of the Integration, Resources, and Effectiveness of Technology in K-12 Classrooms. *J. Inf. Technol. Educ. Res.* 14, (January 2015), 397–416. https://doi.org/10.28945/2298

[15]   Ruiqi Deng, Pierre Benckendorff, and Deanne Gannaway. 2019. Progress and new directions for teaching and learning in MOOCs. *Comput. Educ.* 129, (February 2019), 48–60. https://doi.org/10.1016/j.compedu.2018.10.019

[16]   Mrs Swati Desai and Bharati Vidyapeeth. ROLE OF INFORMATION COMMUNICATION TECHNOLOGIES IN EDUCATION.

[17]   Lu Ding, Tong Li, Shiyan Jiang, and Albert Gapud. 2023. Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *Int. J. Educ. Technol. High. Educ.* 20, 1 (December 2023), 63. https://doi.org/10.1186/s41239-023-00434-1

[18]   Eva Durall and Evangelos Kapros. 2020. Co-design for a Competency Self-assessment Chatbot and Survey in Science Education. . 13–24. https://doi.org/10.1007/978-3-030-50506-6_2

[19]   Amanda R. Ellis and Emily Slade. 2023. A New Era of Learning: Considerations for ChatGPT as a Tool to Enhance Statistics and Data Science Education. *J. Stat. Data Sci. Educ.* 31, 2 (May 2023), 128–133. https://doi.org/10.1080/26939169.2023.2223609

[20]   J. Ferrer-Torregrosa, J. Torralba, M. A. Jimenez, S. García, and J. M. Barcia. 2015. ARBOOK: Development and Assessment of a Tool Based on Augmented Reality for Anatomy. *J. Sci. Educ. Technol.* 24, 1 (February 2015), 119–124. https://doi.org/10.1007/s10956-014-9526-4

[21]   S. Gilles, E. Worden, Z. Luo, S. Sonkar, R. Baranuik, and N. Heffernan. 2024. Utilizing the CLASS Framework to Develop a Conversational AI Tutor for open-ended problems in ASSISTments. (January 2024). Retrieved June 22, 2024 from https://par.nsf.gov/biblio/10470443-utilizing-class-framework-develop-conversational-ai-tutor-open-ended-problems-assistments

[22]   Rich Halverson and R Benjamin Shapiro. 2012. Technologies for education and technologies for learners: How information technologies are (and should be) changing schools. *Wis. Cent. Educ. Res. WCER Work. Pap.* 6, (2012), 33.

[23]   Jonhan Ho, Orly Aridor, and Anil V. Parwani. 2012. Use of contextual inquiry to understand anatomic pathology workflow: Implications for digital pathology adoption. *J. Pathol. Inform.* 3, 1 (2012), 35.

[24]   Karen Holtzblatt, Jessamyn Burns Wendell, and Shelley Wood. 2005. *Rapid contextual design: a how-to guide to key techniques for user-centered design*. Elsevier.

[25]   Miloš Ilić, Vladimir Mikić, Lazar Kopanja, and Boban Vesin. 2023. Intelligent techniques in e-learning: a literature review. *Artif. Intell. Rev.* 56, 12 (December 2023), 14907–14953. https://doi.org/10.1007/s10462-023-10508-1

[26]   Francisco Iniesto, Patrick McAndrew, Shailey Minocha, and Tim Coughlan. 2022. A qualitative study to understand the perspectives of MOOC providers on accessibility. *Australas. J. Educ. Technol.* 38, 1 (2022), 87–101. https://doi.org/10.14742/ajet.6610

[27]   I. C. F. International (Firm). 2015. *Literature review on the impact of digital technology on learning and teaching*. Scottish Government.

[28]   Ma Jinglin. 2020. Research on the Application of Data Mining Technology in Educational Management System. *J. Educ. Pract.* 11, 8 (2020), 108.

[29]   Adrian Johnston. Jacques Lacan (Stanford Encyclopedia of Philosophy). Retrieved July 12, 2024 from https://plato.stanford.edu/entries/lacan/

[30]   Samuel Kai-Wai Chu and David M. Kennedy. 2011. Using online collaborative tools for groups to co-construct knowledge. *Online Inf. Rev.* 35, 4 (2011), 581–597.

[31]   Derviş Kayımbaşıoğlu, Bora Oktekin, and Hüseyin Hacı. 2016. Integration of gamification technology in education. *Procedia Comput. Sci.* 102, (2016), 668–676.

[32]   Margaret C. Keiper, Gil Fried, Joshua Lupinek, and Heidi Nordstrom. 2023. Artificial intelligence in sport management education: Playing the AI game with ChatGPT. *J. Hosp. Leis. Sport Tour. Educ.* 33, (November 2023), 100456. https://doi.org/10.1016/j.jhlste.2023.100456

[33]   Rehan Ahmed Khan, Masood Jawaid, Aymen Rehan Khan, and Madiha Sajjad. 2023. ChatGPT - Reshaping medical education and clinical management. *Pak. J. Med. Sci.* 39, 2 (February 2023). https://doi.org/10.12669/pjms.39.2.7653

[34]   Byung-Hak Kim, Ethan Vizitei, and Varun Ganapathi. 2018. GritNet: Student Performance Prediction with Deep Learning. Retrieved May 19, 2024 from http://arxiv.org/abs/1804.07405

[35]   Sunyoung Kim, Joobo Shim, and Jaechang Shim. 2023. A study on the utilization of OpenAI ChatGPT as a second language learning tool. *J. Multimed. Inf. Syst.* 10, 1 (2023), 79–88.

[36]   Péter Kovács, Niall Murray, Gregor Rozinaj, Yevgeniya Sulema, and Renata Rybarova. 2015. *Application of immersive technologies for education: State of the art*. https://doi.org/10.1109/IMCTL.2015.7359604

[37]   Yanika Kowitlawakul, Jocelyn Jie Min Tan, Siriwan Suebnukarn, Hoang D. Nguyen, Danny Chiang Choon Poo, Joseph Chai, Wenru Wang, and Kamala Devi. 2022. Utilizing educational technology in enhancing undergraduate nursing students' engagement and motivation: a scoping review. *J. Prof. Nurs.* 42, (2022), 262–275.

[38]   Vladimir Kryukov and Alexey Gorin. 2017. Digital technologies as education innovation at universities. (2017).

[39]   Kari L. Kumar and Ron Owston. 2016. Evaluating e-learning accessibility by automated and student-centered methods. *Educ. Technol. Res. Dev.* 64, 2 (April 2016), 263–283. https://doi.org/10.1007/s11423-015-9413-6

[40]   Eugenijus Kurilovas. 2019. Advanced machine learning approaches to personalise learning: learning analytics and decision making. *Behav. Inf. Technol.* 38, 4 (April 2019), 410–421. https://doi.org/10.1080/0144929X.2018.1539517

[41]   Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. *Expert Syst. Appl.* 41, 4, Part 2 (March 2014), 2065–2073. https://doi.org/10.1016/j.eswa.2013.09.005

[42]   William Little and William Little. 2014. *Introduction to Sociology-Chapter 16. Education*. BCcampus. Retrieved June 22, 2024 from https://opentextbc.ca/introductiontosociology/chapter/chapter16-education/

[43]   Bella Martin and Bruce M. Hanington. 2012. *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport Publishers, Beverly, MA.

[44]   Emad Mushtaha, Saleh Abu Dabous, Imad Alsyouf, Amr Ahmed, and Naglaa Raafat Abdraboh. 2022. The challenges and opportunities of online learning and teaching at engineering and theoretical colleges during the pandemic. *Ain Shams Eng. J.* 13, 6 (2022), 101770.

[45]   Cody Nelson, Lisa Hartling, Sandra Campbell, and Anna E. Oswald. 2012. The effects of audience response systems on learning outcomes in health professions education. A BEME systematic review: BEME Guide No. 21. *Med. Teach.* 34, 6 (June 2012), e386–e405. https://doi.org/10.3109/0142159X.2012.680938

[46]   Davy Tsz Kit Ng, Chee Wei Tan, and Jac Ka Lok Leung. 2024. Empowering student self-regulated learning and science education through CHATGPT : A pioneering pilot study. *Br. J. Educ. Technol.* (March 2024), bjet.13454. https://doi.org/10.1111/bjet.13454

[47]   V V Osadchyi, N V Valko, and L V Kuzmich. 2021. Using augmented reality technologies for STEM education organization. *J. Phys. Conf. Ser.* 1840, 1 (March 2021), 012027. https://doi.org/10.1088/1742-6596/1840/1/012027

[48]   T Ross Owen. SELF-DIRECTED LEARNING IN ADULTHOOD: A LITERATURE REVIEW.

[49]   Tung Phung, Victor-Alexandru Pădurean, Anjali Singh, Christopher Brooks, José Cambronero, Sumit Gulwani, Adish Singla, and Gustavo Soares. 2023. Automating Human Tutor-Style Programming Feedback: Leveraging GPT-4 Tutor Model for Hint Generation and GPT-3.5 Student Model for Hint Validation. https://doi.org/10.48550/arXiv.2310.03780

[50]   Martha C. Polson and J. Jeffrey Richardson (Eds.). 1988. *Foundations of Intelligent Tutoring Systems*. Psychology Press, New York. https://doi.org/10.4324/9780203761557

[51]   Ava G. Porter and Stuart Tousman. 2010. Evaluating the Effect of Interactive Audience Response Systems on the Perceived Learning Experience of Nursing Students. *J. Nurs. Educ.* 49, 9 (September 2010), 523–527. https://doi.org/10.3928/01484834-20100524-10

[52]   Plews Rachel C. 2017. SELF-DIRECTION IN ONLINE LEARNING: THE STUDENT EXPERIENCE. *Int. J. Self-Dir. Learn.* 14, (Spring 2017).

[53]   Awaz Naaman Saleem, Narmin Mohammed Noori, and Fezile Ozdamli. 2022. Gamification Applications in E-learning: A Literature Review. *Technol. Knowl. Learn.* 27, 1 (March 2022), 139–159. https://doi.org/10.1007/s10758-020-09487-x

[54]   Mageswaran Sanmugam, Norasykin Mohd Zaid, Hasnah Mohamed, Zaleha Abdullah, Baharuddin Aris, and Salihuddin Md Suhadi. 2015. Gamification as an educational technology tool in engaging and motivating students; An analyses review. *Adv. Sci. Lett.* 21, 10 (2015), 3337–3341.

[55]   Olga C. Santos, Jesús G. Boticario, Alicia Fernández Del Viso, Santiago Pérez De La Cámara, Carlos Rebate Sánchez, and Emmanuelle Gutiérrez Y Restrepo. 2007. Basic Skills Training to Disabled and Adult Learners Through an Accessible e-Learning Platform. In *Universal Access in Human-Computer Interaction. Applications and Services*, Constantine Stephanidis (ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 796–805. https://doi.org/10.1007/978-3-540-73283-9_86

[56]   Teodor Savov, Valentina Terzieva, and Katia Todorova. 2018. Computer Vision and Internet of Things: Attention System in Educational Context. In *Proceedings of the 19th International Conference on Computer Systems and Technologies* (*CompSysTech '18*), 2018. Association for Computing Machinery, New York, NY, USA, 171–177. https://doi.org/10.1145/3274005.3274014

[57]   Johannes Schneider, Bernd Schenk, and Christina Niklaus. 2024. Towards LLM-Based Autograding for Short Textual Answers: In *Proceedings of the 16th International Conference on Computer Supported Education*, 2024. SCITEPRESS - Science and Technology Publications, Angers, France, 280–288. https://doi.org/10.5220/0012552200003693

[58]   Hongxia Sun, Ahmed Farouk, and Dou Zhen. 2019. Study on application of data mining technology in university computer network educational administration management system. *J. Intell. Fuzzy Syst.* 37, 3 (September 2019), 3311–3318. https://doi.org/10.3233/JIFS-179133

[59]   Viriya Taecharungroj. 2023. "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *Big Data Cogn. Comput.* 7, 1 (February 2023), 35. https://doi.org/10.3390/bdcc7010035

[60]   Yu-Ching Tseng and Yi-Hsuan Lin. 2024. Enhancing English as a Foreign Language (EFL) Learners' Writing with ChatGPT: A University-Level Course Design. *Electron. J. E-Learn.* 22, 2 (April 2024), 78–97. https://doi.org/10.34190/ejel.21.5.3329

[61]   Faiz Tuma. 2021. The use of educational technology for interactive teaching in lectures. *Ann. Med. Surg.* 62, (2021), 231–235.

[62]   Herman G. Van De Werfhorst and Jonathan J.B. Mijs. 2010. Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective. *Annu. Rev. Sociol.* 36, 1 (June 2010), 407–428. https://doi.org/10.1146/annurev.soc.012809.102538

[63]   Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30, (2017).

[64]   Ya-huei Wang and Hung-Chang Liao. 2011. Data mining for adaptive learning in a TESL-based e-learning system. *Expert Syst. Appl.* 38, 6 (June 2011), 6480–6485. https://doi.org/10.1016/j.eswa.2010.11.098

[65]   Françeska Xhakaj, Vincent Aleven, and Bruce M. McLaren. 2016. How Teachers Use Data to Help Students Learn: Contextual Inquiry for the Design of a Dashboard. In *Adaptive and Adaptable Learning*, Katrien Verbert, Mike Sharples and Tomaž Klobučar (eds.). Springer International Publishing, Cham, 340–354. https://doi.org/10.1007/978-3-319-45153-4_26

[66]   Yangyu Xiao and Yuying Zhi. 2023. An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages* 8, 3 (2023), 212.

[67]   Da Yan. 2023. Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Educ. Inf. Technol.* 28, 11 (November 2023), 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

[68]   Xiaohua Yu, Jueqi Guan, and Jing Leng. 2016. Using Learning Analytics to Support Personalized Learning and Quality Education: A Case Study of China's "Everyone Connected" Project. In *Current Developments in Web Based Learning*, Zhiguo Gong, Dickson K. W. Chiu and Di Zou (eds.). Springer International Publishing, Cham, 196–201. https://doi.org/10.1007/978-3-319-32865-2_21

[69]   2023. The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *J. Appl. Learn. Teach.* 6, 1 (May 2023). https://doi.org/10.37074/jalt.2023.6.1.29

[70]   Khan Academy | Free Online Courses, Lessons & Practice. *Khan Academy*. Retrieved June 7, 2024 from https://www.khanacademy.org/

[71]   Coursera | Online Courses From Top Universities. Join for Free. Retrieved June 7, 2024 from https://www.coursera.org/

[72]   ChatGPT. Retrieved June 22, 2024 from https://chatgpt.com

[73]   Meet Khanmigo: Khan Academy's AI-powered teaching assistant & tutor. Retrieved June 7, 2024 from https://khanmigo.ai/

[74]   Interactive visual computer vision analysis based on artificial intelligence technology in intelligent education | Neural Computing and Applications. Retrieved June 8, 2024 from https://link.springer.com/article/10.1007/s00521-021-06285-z

[75]   Khanmigo for learners: Always-available tutor, powered by AI. Retrieved June 22, 2024 from https://khanmigo.ai/learners

[76]   Podium.page - Try for FREE. *Podium*. Retrieved June 22, 2024 from https://podium.page

[77]   Miro. Retrieved June 22, 2024 from https://miro.com/app/dashboard/

[78]   AutoGen | AutoGen. Retrieved July 12, 2024 from https://microsoft.github.io/autogen/

[79]   Hello GPT-4o. Retrieved July 13, 2024 from https://openai.com/index/hello-gpt-4o/

[80]   Stanford Encyclopedia of Philosophy. Retrieved July 13, 2024 from https://plato.stanford.edu/

[81]   Tavily. Retrieved July 10, 2024 from https://tavily.com/

[82]   Tutor AI - Your personal AI tutor to learn anything. *tutorai.me*. Retrieved July 10, 2024 from https://tutorai.me