



# Designing Conversational Agents to Support Student Teacher Learning in Virtual Reality Simulation: A Case Study

Chih-Pu Dai\*

Department of Learning Design and Technology, University of Hawai'i at Mānoa, United States  
cdai@hawaii.edu

Alex Barrett

Department of Educational Psychology and Learning Systems, Florida State University, United States  
abarrett3@fsu.edu

Sherry A. Southerland

School of Teacher Education, College of Education, Florida State University United States  
ssoutherland@fsu.edu

Fengfeng Ke

Department of Educational Psychology and Learning Systems, Florida State University, United States  
fke@fsu.edu

Luke West

Department of Educational Psychology and Learning Systems, Florida State University, United States  
lawest@fsu.edu

Nuodi Zhang

Department of Educational Psychology and Learning Systems, Florida State University, United States  
nzhang4@fsu.edu

Saptarshi Bhowmik

Department of Computer Science, Florida State University, United States  
sbhowmik@fsu.edu

Xin Yuan

Department of Computer Science, Florida State University, United States  
xyuan@cs.fsu.edu

## ABSTRACT

Maximizing educational impacts with learning technologies is one of the areas that researchers and practitioners are concerned about in the field of Human-Computer Interaction (HCI) and human-centered artificial intelligence (HCAI). In this case study, we report user experiences and lessons learned of the Enactive Virtual Environment for teaching practice (*EVETeach*) with AI-powered virtual student agents called *Evelyn*. We conducted a user study with a case study research design. We collected multiple sources of data from 24 student teachers, including participatory observations, field notes, semi-structured interviews, computer-based conversation logs, audio-, video-, and screen-recordings, and a cognitive walkthrough. We identified the following salient emerging findings as lessons learned: 1) Student teachers value and relate to the teaching practices in virtual reality simulation with AI-powered conversational agents, 2) AI-powered conversational agents inject humor to facilitate situational and social teaching practice, and 3) AI-powered conversational student agents maintain authentic discourse to promote student teachers' pedagogical reasoning.

## CCS CONCEPTS

- Interactive systems and tools; • Empirical studies in HCI; • Emerging technologies;

\*Corresponding author

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0331-7/24/05

<https://doi.org/10.1145/3613905.3637145>

## KEYWORDS

Virtual reality, Artificial intelligence, Teacher education, Conversational agents

## ACM Reference Format:

Chih-Pu Dai, Fengfeng Ke, Nuodi Zhang, Alex Barrett, Luke West, Saptarshi Bhowmik, Sherry A. Southerland, and Xin Yuan. 2024. Designing Conversational Agents to Support Student Teacher Learning in Virtual Reality Simulation: A Case Study. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3613905.3637145>

## 1 INTRODUCTION

Optimizing educational impacts with learning technologies is one of the areas that researchers and practitioners are concerned about in the field of Human-Computer Interaction (HCI) [1, 14]. Strategically improving student teachers' teaching capabilities with learning technologies is one aspect of optimizing educational impacts because it has been argued that the quality of student teachers determines the educational impacts [14, 19]. Hence, preparing student teachers is critical to foster positive outcomes for student learning and broader influences. There are many challenges for student teacher preparation, including the lack of practice opportunities for teaching [19]. Leveraging learning technologies can potentially address such insufficiency. For example, virtual reality (VR) simulation has been introduced in teacher education to increase practice opportunities [24]. VR simulation creates spaces for student teachers to practice teaching and approximate veteran teachers' professional practices by sensing and reflecting their own practices *in situ* [21]. In other words, reflective teaching practices are crucial in clinical teaching and hence the interactions in the VR classroom simulations with virtual conversational agents (CAs) are key to facilitating such reflective stance [13, 21]. The authenticity of interactions occurring in VR simulation is determined by several

factors, such as locus of control, scenario designs, and CA [21]. In particular, CAs can be designed to afford authentic interactions and shape classroom practice discourse. In such practice, student teachers can engage in decision making that entails by CAs. Integrating artificial intelligence (AI) into such learning technologies may augment the capabilities of the learning technologies and increase interactivity [16] thereby promoting teacher learning. However, student teachers' interactions with AI CAs in VR simulation have not been thoroughly explored and the implications for the field of HCI have been inadequate in this context. The purpose of this current case study is to explore heuristics in relation to the interactions between student teachers and AI CAs and to share lessons learned from the HCI perspectives.

Specifically, we investigated the following research questions (RQs): RQ1: What were the student teachers' experiences of teaching practices in the designed VR simulation with AI-powered CAs? RQ2: How does the designed VR simulation with AI-powered CAs afford unique spaces for student teachers' teaching knowledge and skills development?

## 2 RELATED WORK

### 2.1 Human-centered design for AI-powered CAs in VR simulation

CAs designs varied in different contexts, such as business, health industry, entertainment and gaming, or education. In the current study, we sought to explore the designs of AI in CAs with value-sensitive design [3] and a human-centered design approach [3, 4, 7]. By probing user needs, experiences, usability, and potential limitations, the integration of value-sensitive design [3] and a human-centered design approach [3, 4, 7] emphasizes human-centeredness, ensuring responsible and impactful product design for the public good. Both frameworks aim to ground designers in designing technologies with high accessibility and societal impacts [3, 4]. Our societal context lies in the educational impact as a result of student teacher learning.

Similarly, AI-integrated CAs have found applications in diverse areas such as business [11], health care [12], and education [13]. Scholars have argued that CAs with AI have been intricately blended in VR simulation [20]. In the current participatory design-based research project, we situate in value-sensitive design [3] and human-centered design approach [3, 4, 7] to conceptualize how student teachers interact with AI CAs in VR simulation. Specifically, derived from value-sensitive design and human-centered design [3, 4], the study of human-centered artificial intelligence (HCAI) has emerged [5, 15]. Capel and Brereton posited that HCAI considers the characteristics, experiences, and perspectives of human users, including their values and agency [5].

### 2.2 Humor in HCAI and CAs

Researchers and practitioners in HCI, or HCAI, are interested in investigating the role that humor plays in CAs as it could foster socially oriented interactions, improve user experience, and increase motivation and efforts [2, 8]. Research in social psychology and HCI has contributed to the field with regard to the forms of humor: *affiliative*, characterized by the use of gentle humor to strengthen

social connections; *self-defeating*, where humor is employed to bolster relationships at one's own cost; *self-enhancing*, which involves using humor to boost one's representations; and *aggressive*, which entails using humor to elevate oneself while belittling others [23]. Humor can also be enacted in different ways, including riddles and punning or corny jokes, and dry humor (e.g., witty remarks, irony, sarcasm, or clever play on words) [2].

Empirical studies on humor and HCAI are limited. Humor was one of the characteristics that emerged from Veletsianos et al.'s phenomenological study [22]. Veletsianos et al.'s study was one of the earlier explorations that identified humor as one feature in human-agent interactions in an education context [22], but not in a teacher learning context. More recently, Ceha et al. [2] designed CAs with three different forms of humor enactment, that is, *an affiliative style* (i.e., telling jokes and being funny to amuse others for interpersonal relationship building), *a self-defeating style* (i.e., amusing other at one's own cost or by self-deprecating), or *no humor*. With an experimental design study, they found that CAs (as teachable agents) designed with an affiliative style of humor can enhance motivation and efforts, while the effects on learning were not significantly different between three conditions. Designing AI-powered virtual agents to help users practice interviews, Zhou and colleagues revealed that AI-powered virtual agents designed to be *humorous* are more beneficial regarding social interactions. For example, the AI-powered virtual agent designed to be humorous was rated to be more likable, helpful, trustworthy, and insightful than the one that was not designed with humor [10].

The functions of computational humor [8] in student teacher's simulated clinical practices can be regarded as a tool to provide social cues for student teachers to practice pedagogical reasoning and decision making, including classroom management. It can also foster social relationships and affect student teachers' emotions. Studying and appraising computational humor is situational in verbal events and conversations [9]. In our teacher education context with AI-powered CAs, perceived computational humor could represent students' creativity and imagination. It is worthwhile to further explore humor in CAs as to how it can be leveraged to facilitate teacher learning.

### 2.3 Teacher learning in VR simulation

VR simulation has been used for teacher learning with affordances such as virtual role-play [26], practice-based learning [21], and broader opportunity for simulated teaching practicum [13]. VR simulation has been found to be valuable in improving student teachers' noticing of students' in-class behaviors as well as student teachers' exercise of domain specific knowledge and pedagogical knowledge in teaching activities [21].

However, there are several challenges facing teacher learning in VR simulation. For example, several teacher learning platforms in VR simulation have integrated behind-the-scene human actors to interact with end users; cf. [27]. In doing so, the design may undermine the case for VR simulations as viable alternative platforms for teacher learning, compromising the cost-efficiency and convenience that make VR simulation a valuable tool for preservice student teachers to practice anytime [27].

### 3 STUDY DESIGN

To explore how student teacher users interact with the system and the unique patterns resulting in the interactions between student teachers and AI technologies designed in virtual CA students, we conducted a user study with a case study research design [28]. Case study is suitable for our purposes to highlight the user experience and lessons learned for value-sensitive and human-centered design [3, 4, 7]. We reported the first iteration of the user study and emerging findings in this paper.

#### 3.1 The design of CA with AI in VR for student teachers

The Enactive Virtual Environment for teaching practice (*EVETeach*) [24] is an ongoing design-based research project that was designed and developed to support student teacher learning with a broader goal of contributing to societal impacts on teacher education. Toward this goal, *EVETeach* was created in *OpenSimulator* (or *OpenSim*), an open-source, customizable, and scalable VR space for learning and making [24].

In *EVETeach*, CAs with AI (called *Evelyn*) were designed and implemented in the VR simulation using *OpenSim*. Figure 1 shows the workflow of the design and development of *EVETeach* [24]. *Evelyn*—the virtual CAs that simulate human students in secondary education in math and science were powered by generative AI (i.e., GPT-2 from *OpenAI*) [25]. We evaluated various generative AI models, considering factors such as their open-source nature and the potential for customization to align with our specific VR experiences [31]. Our design-based decision was that GPT-2 was the best fit for our exploratory purposes. In the experimental classroom scenario, six CAs were deployed. The student teacher can interact with an individual CA student or address the entire class within a dynamic setup of a U-shaped classroom configuration. Within this exploratory stage, the student teacher interacted with the CAs through text-based input in *EVETeach* and the output of the CAs was stored in the VR simulation (i.e., a dialog log window in the user interface in the *OpenSim* VR simulation), allowing student teachers to review the record of the conversation during the in-the-moment teaching practices. Figure 2 demonstrates an example of the classroom configuration. Figure 3 shows the dialog log window.

#### 3.2 Setting, participants, and procedure

The research team engaged in participatory design and development of *EVETeach* as well as recruited student teachers in the southeastern US. Over the past three years, the research team iteratively designed and developed *EVETeach* with *Evelyn* [24]. The research team comprised members from disciplines of learning technologies, teacher education, and computer science.

Using participatory design and user testing approaches, we collected data with student teachers ( $n = 24$ ) from math and science disciplines. Data were collected from participatory observations, field notes, semi-structured interviews, computer-based conversation logs, audio-, video-, and screen-recordings, and think-aloud as a cognitive walkthrough [17–18]. The participatory observations and semi-structured interviews focused on HCAI and teaching practices with AI-powered CAs in VR for the design implications and lessons learned. During the teaching practices, student teachers

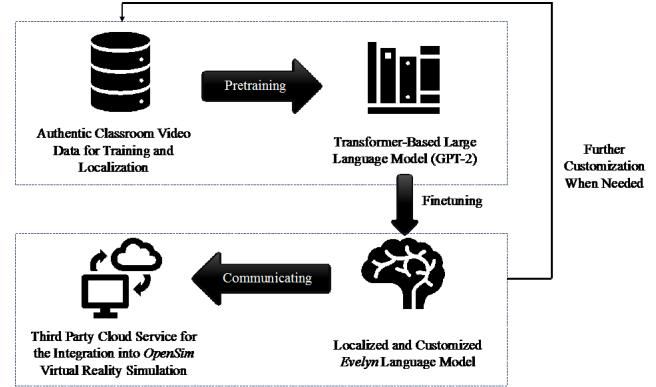


Figure 1: Workflow of the design and development of *Evelyn* and *EVETeach*, a product of [24].



Figure 2: Classroom scenario configuration of *Evelyn* and *EVETeach*, a product of [24].



Figure 3: Dialog log window is a feature in *Evelyn* and *EVETeach*, a product of [24].

were encouraged to verbalize what they were thinking and doing in the VR simulation.

The student teachers were interviewed to gather their insights and experiences regarding AI-powered CA students, focusing on

aspects such as *humanlike*, *realistic*, and *engaging* [29] (the student teacher participants were asked to rate these aspects from 1 to 5, one being strongly disagree; five being strongly agree, before and after the experience), they were also asked to provide qualitative evaluations on the usage of AI-powered CAs in teacher education. All names are pseudonyms in this case study.

In the case study, the participating student teachers signed up voluntarily for a two-hour session with a facilitator. In the naturalistic study setting, they can choose to practice either math, gas law, or fungi in secondary math and science curricula. At the beginning of the session, they were briefed about the study session, *EVETeach* and *Evelyn* [24], they were told that the facilitator was only in the session to provide technical support and to facilitate the session, as well as to conduct data collection; they were informed that their participation was completely voluntary and confidential. We focused our analysis on their interactions with and pedagogical affordances of the AI CAs. We constantly reflected on the data, comparing information across participants, sessions, and multiple different data sources. This process involved constant comparison analysis techniques [30], applying the researchers' interpretive lens. We used open coding for preliminary emerging findings.

## 4 FINDINGS

In this section, we present our findings on the lessons learned and the student teachers' experience in *EVETeach*. Three relevant and salient emerging themes developed from the qualitative data addressed our RQs. The three themes depicted users' value for and experiences within the designed learning environment (Section 4.1 for RQ1) and the affordances of AI-powered CA students, including the unique social features (Section 4.2 for RQ2) and pedagogical attributes for teacher education (Section 4.3 for RQ2). These themes derived from naturalistic inquiry, through constant comparison [30], provided us with the insights and lessons regarding the use of AI-powered CA students, laying the groundwork for further implications.

### 4.1 Student teachers value and relate to the teaching practices in virtual reality simulation with AI-powered CAs

*Human agency* in making decisions is critical in the modern HCAI paradigm. To this end, we derived the lessons learned related to the *EVETeach* and *Evelyn* from the perspectives of student teachers and their agency in empowering values and identifying teacher-agent interaction features for the learning technologies designed. Student teachers reasoned about the educational impacts of *EVETeach* during and after the learning experience.

A frequent question participating student teachers asked was: “*Are there humans behind them?*” This question asked by the participants indicated that *Evelyn*'s interactions in *EVETeach* were authentic, with high fidelity. Qualitative ratings of the HCI aspects of the AI-powered CAs revealed that the agents designed were realistic, Ivory (P04) gave a rating of two for *realistic* in the pre-experience interview and a four in the post-experience interview, she commented: “*they look like [a] classroom...and responses that kids would give.*” In terms of *humanlike*, she gave a rating of two in the pre-session interview and between three and four after the

experience, she explained: “*it couldn't be five because they are not human.*” Charlotte (P06) gave similar ratings, she rated four for *realistic* in the pre-experience interview and five in the post-experience interview, stating: “*like the whole interactive part was realistic.*” For *engaging*, she gave a rating of four, pre-experience, and five, post-experience, claiming: “*they all asked a lot of questions and they had a lot of input for everything that I said.*” In sum, Charlotte (P06) commented: “*I think overall they all have great personalities, and I feel like they were just very engaging, and fun to teach, and they have funny comments [smiling].*”

Laura (P09) rated four for *engaging* and three for *realistic* in the post-experience interview, she supported her rating for the *realistic* category by saying: “*real students would elaborate more.*” She further explicated the educational impacts of *EVETeach* and *Evelyn* in teacher education: “*I think it would really help preservice teachers of all levels, like I said, practice the talk moves, in a very explicit way and force them, like I said, to plan ahead in terms of the goals of the conversations...and the science behind the talk, how does the science work, what's the phenomenon, what do I want my students to get out of this. I think these are beneficial for preservice teachers.*”

### 4.2 AI-powered CAs inject humor to facilitate situational and social teaching practice

Relating to Charlotte's (P06) comments that AI-powered CAs “*have funny comments*” above, another finding in this case study pertaining to the use of humor by the AI-powered CAs in the simulated classroom interactions to accommodate socially oriented pedagogical reasoning and classroom management. An important aspect of the computational humor in HCAI is that it should be recognizable and interpretable to the end users (i.e., student teachers) so that they may address the humor (i.e., to respond or not to respond) and extend the social agency and dynamicity in the human-AI interactions.

At times, preservice teachers would interpret AI-powered CA students' response as a “*joke*.” In other words, AI-powered CA students created a practice environment with social cues for the student teachers. In the following examples (Table 1), Hazel (P14) has tried to engage students in discussing science, but the AI-powered CA student (named John) was not engaged, as she interpreted student's responses as jokes.

Another type of humor the AI-powered CAs demonstrated is *self-enhancing* humor. *Evelyn* was observed to deliver discourse that would be considered as humor that can improve one's representation (see the following excerpt in Table 2). However, in this instance, Emma (P03) didn't seem to respond to the humorous delivery by the AI-powered CA, instead she delivered scientific information to the class, a pedagogical decision made to focus on disciplinary practices, leading to a “*silent*” response.

In a classroom environment, the humor exhibited by AI-powered CA students can be viewed as both a pedagogical challenge and an opportunity for student teachers to assess and formulate age-appropriate responses. *Evelyn* was specifically designed to emulate students in secondary education, who typically fall within the young adult category with a fair grasp of real-world understanding.

**Table 1: Hazel (P14)**

Event Line	Interaction Context	Excerpt
1	Hazel to AI-powered CAs	They must absorb or get nourishment from other sources. What are some of those sources?
2	AI-powered CA (John)	Uh, plants.
3	Hazel to AI-powered CA	Okay, John, that's good!
	Hazel think-aloud	Someone else... [being interrupted as she noticed AI CA's animation; but she submitted this input to the AI-powered CAs anyway]
4	Hazel think-aloud	Did somebody raise their hands (after checking students in VR)? No, they rested their arms, sorry.
5	AI-powered CA (Thomas)	Mister Ormsby?
6	Hazel think-aloud	[Read out loud] Mister Ormsby? [laughing] Okay, <i>he's making jokes? Is that Thomas making jokes</i> [laughing]?
7	Facilitator to Hazel	Yeah, why do you think that it's a joke?
8	Hazel to facilitator	Cause they are probably not interested right now, talking about fungi, of course, (talking while laughing) you know how students do. I guess I should say to them "it's not like a joke [raising tone]!"
9	Hazel to AI-powered CAs	Listen everyone, it's not a joke. Let's be serious.
10	AI-powered CAs (Linda)	I am gonna be successful with this.
11	Hazel to AI-powered CAs	Yes, you are Linda. Studying fungi can be pretty interesting.
12	AI-powered CAs (Karen)	How did fungi acquire nutrients?

**Table 2: Emma (P03)**

Event Line	Interaction Context	Excerpt
1	Emma to AI-powered CA	Good job, Thomas.
2	AI-powered CA (Thomas)	I know.
3	Emma to AI-powered CAs (class)	Fungi multiply through spores
4	AI-powered CA	[silent]

**Table 3: Emily (P05)**

Event Line	Interaction Context	Excerpt
1	Emily to AI-powered CAs	If eaten, the poison of the mushroom can negatively affect the body or the organs of whatever ate it.
2	AI-powered CA (Michael)	Ok, so let's take this imaginary mushroom and then I swallow it.
3	Emily to AI-powered CAs	That's funny. But don't go looking for mushrooms now.

Among this age group, it's not uncommon for some students to engage in humor involving unusual topics or playful statements with imagination. Following is an example of Emily and an AI-powered CA student (Table 3).

### 4.3 AI-powered CA students maintain authentic discourse to promote student teachers' pedagogical reasoning

Student teachers' engagement in pedagogical reasoning was also found to be an emerging phenomenon supported by AI-powered CA students. In a teaching practice event, Nicole (P10), exercised pedagogical reasoning evidenced by her revision of thinking in

multiple rounds of the interaction with AI-powered CA students. In her first attempt, she sought to assess students' learning progress by initiating with the phrase, "So, let's review." Applying a think-aloud protocol, she remarked, "Let's see if anyone responds to that." The AI-powered students were not ready to participate, as reflected by "[Silent]," attributed to a student named Michael. In the next round of interaction (Table 4), Nicole (P10) learned, from the previous interactions with the AI-powered CA students, that the input she provided to the students should be more granular in details, otherwise students would be confused.

Similarly, Madeline (P01) adopted rephrasing, as a pedagogical reasoning strategy, in addressing AI-powered CA student's ("Thomas" in the following excerpt, see Table 5) learning performance to help

**Table 4: Another round of interactions from Nicole (P10)**

Event Line	Interaction Context	Excerpt
1	Nicole to AI-powered CA	What happens when temperature increases?
2	Nicole think-aloud to the facilitator	I should be more specific.
3	Nicole to AI-powered CA	What happens to molecules when temperature increases?
4	AI-powered CA (Karen)	They shrink
5	Nicole think-aloud to the facilitator	[Molecules shrink? What does that even mean? I don't think so...]
6	Nicole to AI-powered CA and think-aloud to the facilitator in the bracket	Well, they don't quite shrink [they don't shrink? I don't think they shrink?] but they do move around faster, because of the higher pressure
7	Nicole think-aloud to the facilitator	Actually not because of the higher pressure, which causes the higher pressure, well [head tilted indicating thinking] I am not sure about the cause-effect situation

focus the interactions, and ultimately reached her microteaching objective.

## 5 DISCUSSION, FUTURE WORK, AND LIMITATIONS

In this case study, we derived valuable heuristics from student teachers for the design and development of AI-powered CAs in VR simulation, aimed at enhancing the learning experience of student teachers and fostering broader educational impacts in HCAI [3, 4, 7].

Overall, student teachers regarded positively for *Evelyn* in terms of *humanlike*, *realistic*, and *engaging*. Meanwhile, the findings also suggested that *Evelyn* can be further improved in terms of the variety of conversations they can carry on. From the perspectives of human agency interacting with the technology, student teachers value the affordances of *EVETeach* and advised that *EVETeach* can help student teachers to practice discourse-oriented teaching in math and science classrooms.

Computational humor has been approached through the lenses of social psychology and HCI. Expanding upon prior classifications of computational humor [2, 23], we proficiently identified and interpreted how student teachers construe humor expressed by AI-powered student agents. Further, previous research has found that AI-powered CAs tended to demonstrate “dark” humor [22]. We add to the literature and maintain that humor exhibited by AI-powered student agents within a teacher learning context presents an opportunity for student teachers to enhance their pedagogical decision-making and noticing skills while effectively managing classrooms to nurture productive scientific reasoning for students. Future research in HCAI can expand in this rich area of research by leveraging eye gaze, facial recognition, and humor-related algorithms [8] to explore how humor in HCAI can play a role in student teachers’ psychological states and self-efficacy in teaching.

An earlier issue reported in the literature suggests that automated CA students are not compelling to the student teachers [6]. In our design, we put forward that through an iterative revision of thinking and teaching practices with purposeful reflections, student teachers can reach their microteaching pedagogical objectives by interacting with AI-powered CA students. In other words, the conversations and meaning negotiations afforded by AI-powered

CA students have led student teachers (i.e., human end users) to refine their teaching practices and discourses used to help students for science reasoning.

Our user study with a case study research design provided *three lessons learned and implications* for the fields of HCAI and teacher education. *First*, our findings in naturalistic settings with user experiences highlighted the design aspects of AI-powered CAs (humanlike, realistic, and engaging) suggesting that the AI-powered CA students in the VR simulation were designed with high fidelity. The VR simulation was also valued and perceived by the participants as a promising platform for fostering constructive teacher learning (see Section 4.1). *Second*, the naturalistic interactions between student teachers and AI-powered CAs have brought to light interesting prospects of computational humor studies in HCAI in the teacher education context. Preliminarily, computational humor in HCAI in a teacher education context may provide social cues and may be used as another mechanism for cultivating teaching skills. *Finally*, large language model-integrated AI-powered CA students have been found to afford productive teaching practices through real-time conversations that prompted reflective teaching. Although our case study provided lessons learned and implications, one limitation is its exploratory nature. Another limitation is the use of AI technologies, that is, GPT-2, which might not reflect the latest advancements in the field of AI. Finally, our findings are a reflection of our worldview and interpretation. We established transferability through constant comparison, researcher reflexivity, and the description of contextual information. However, it is essential to acknowledge the inherent subjectivity associated with these efforts, which should be taken into account when interpreting the findings of this study. Nonetheless, our findings may offer useful insights for further applications and studies of HCAI and education.

## ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. National Science Foundation under Grant 2110777. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. National Science Foundation.

**Table 5: Madeline (P01)**

Event Line	Interaction Context	Excerpt
1	The facilitator to Madeline	What did Thomas say...
2	Madeline to the facilitator	He said air caused the crush
3	The facilitator to Madeline	Is it totally wrong?
4	Madeline to the facilitator	No. It's not totally wrong, it was just a little... like <i>I want him to elaborate on it so I can see that he is thinking</i> ...cause it didn't...it was like a short answer.
5	The facilitator to Madeline	Okay.
6	Madeline to Thomas	What do you mean?
7	Thomas	Like it's hard for the enzymes to break down, and it's not like it needs to be broken down.
8	Madeline think-aloud	No...[in frustration, suggesting that Madeline regarded Thomas's response as an alternative, off-topic one]
9	Madeline think-aloud	[Ask him again, if he kept doing this [being off-topic], I'll ask the whole class.]
10	Madeline to Thomas	What do you mean?
11	Thomas	Like, like, all the molecules are moving around.
12	Madeline think-aloud	Um. ....[Madeline engaged in deep thinking]. Let's see [talking to Thomas again].
13	Madeline to Thomas	Why do you think air crushed the tanker car? [Re-read her input to Thomas carefully by selecting/highlighting the whole sentence in the user interface in <i>OpenSim</i> , and then sending it to Thomas, see Figure 2, the blue dialog box at upper-right as an example].
14	Thomas	Because it compressed and then it got denser.
15	The facilitator to Madeline	I noticed that you changed your question from "what do you mean" to "why do you think air caused the tanker car to crash."
16	Madeline to the facilitator	Um-hum
17	The facilitator to Madeline	What contributed to that change?
18	Madeline to the facilitator	I didn't want him to think that I was asking about the polymer chain [an alternative topic]. So, he has just said so like... I asked him "What do you think crushed the tanker car, and he said "air," so <i>when I just said "what do you mean?" He seemed to get a little off-topic, so I kind of used...sort of...like a guiding question</i> , I guess, I just restated what he said, so...instead of "what do you mean," I used "why do you think that?"
19	The facilitator to Madeline	Yeah. So, you made a more open-ended question to a more...focused question?
20	Madeline to the facilitator	Right. So that I think he would stop talking about [laughing] polymer chains.
21	The facilitator to Madeline	Do you think that helps? And do you think that would be a strategy you would use...in your...classroom teaching?
22	Madeline to the facilitator	I think that it helped in this case...um... and <i>I definitely think that it can help...guide students' thinking</i> ...um...but <i>I think that's something that you need to be careful of doing... in a classroom</i> ...so you do want to guide students' thinking, but you don't want to like force them on this narrow path. So, a broad question can be good to build discussion, but if you just want them to respond to a specific, then I guess...yeah...I would say narrow question like this will do that.
23	Facilitator's observations and Madeline's question to Thomas	Madeline opened the dialog log window in <i>OpenSim</i> VR simulation, she reviewed the conversations, and then asked Thomas: "What got compressed?"
24	Thomas	The steam.
25	Madeline think-aloud	Okay, I think that's the answer.
26	Madeline think-aloud	So now I think I would start like building the classroom consensus from here.

## REFERENCES

[1] Ibrahim, S., Vasalou, A., & Benton, L. 2022. Understanding the situated practices of school technology leaders in the early stages of educational technology adoption. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*, ACM, New York, NY, USA, 1-14.

[2] Ceha, J., Lee, K. J., Nilsen, E., Goh, J., & Law, E. 2021. Can a Humorous Conversational Agent Enhance Learning Experience and Outcomes? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*, ACM, New York, NY, USA, 1-14.

[3] Friedman, B. 1996. Value-sensitive design. *Interactions* 3, 6 (Nov. 1996), 16–23. <https://doi.org/10.1145/242485.242493>

[4] Harrison, S., Tatar, D., & Sengers, P. 2007. The three paradigms of HCI. In *Alt. Chi. Session at the SIGCHI Conference on human factors in computing systems* San Jose, California, USA (CHI '07 Extended Abstracts), 1-18.

[5] Capel, T., & Brereton, M. 2023. What is Human-Centered about Human-Centered AI? A Map of the Research Landscape. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, 1-23.

[6] Ade-Ojo, G. O., Markowski, M., Essex, R., Stiell, M., & Jameson, J. 2022. A systematic scoping review and textual narrative synthesis of physical and mixed-reality simulation in pre-service teacher training. *Journal of Computer Assisted Learning* 38, 3 (May 2022), 861–874. <https://doi.org/10.1111/jcal.12653>

[7] Auernhammer, J. 2020. Human-centered AI: The role of Human-centered Design Research in the development of AI. In Boess, S., Cheung, M., and Cain, R. (Eds.), *Synergy - DRS International Conference 2020*, August 11-14, Held online. <https://>

//doi.org/10.21606/drs.2020.282

[8] Hulstijn, J., & Nijholt, A. (Eds.). 1996. Computational Humor. In *IWCH'96*, University of Twente, 1996.

[9] Anton Nijholt, Oliviero Stock, Alan Dix, and John Morkes. 2003. Humor modeling in the interface. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. Association for Computing Machinery, New York, NY, USA, 1050–1051. https://doi.org/10.1145/765891.766143

[10] Zhou, M. X., Mark, G., Li, J., & Yang, H. 2019. Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 9, 2-3 (June 2019), 1–36. https://doi.org/10.1145/3324774

[11] Rafael Bavaresco, Diego Silveira, Emmanuel Reis, Juliano Barbosa, Renato Righi, Carlos Costa, and Cláudio Moreira. 2020. Conversational agents in business: A systematic literature review and future research directions. *Computer Science Review* 36 (2020), 100239. https://doi.org/10.1016/j.cosrev.2020.100239

[12] Lucia Tudor Car, Dinh-Le Dhinagaran, Bhone Myint Kyaw, Tilman Kowatsch, Shafiq Joty, Yin Leng Theng, and Rifaq Atun. 2020. Conversational agents in health care: scoping review and conceptual analysis. *Journal of Medical Internet Research* 22, 8 (2020), e17158. https://doi.org/10.2196/17158

[13] Ke, F., Dai, Z., Dai, C. P., Pachman, M., Chaulagain, R. S., & Yuan, X. (2020). Designing Virtual Agents for simulation-based learning in Virtual Reality. In *Cognitive and affective perspectives on immersive technology in education* (pp. 151-170). IGI Global. https://doi.org/10.4018/978-1-7998-3250-8.ch008

[14] Stephanidis, C., Salvendy, G., Antona, M., Chen, J. Y., Dong, J., Duffy, V. G., ... & Zhou, J. (2019). Seven HCI grand challenges. *International Journal of Human-Computer Interaction*, 35(14), 1229–1269. https://doi.org/10.1080/10447318.2019.1619259

[15] Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., ... & Xu, W. (2023). Six human-centered artificial intelligence grand challenges. *International Journal of Human-Computer Interaction*, 39(3), 391-437. https://doi.org/10.1080/10447318.2022.2153320

[16] Dai, C. P., & Ke, F. 2022. Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review. *Computers and Education: Artificial Intelligence*, 3, 100087. https://doi.org/10.1016/j.caiei.2022.100087

[17] Wharton, C. W., Reiman, J., Lewis, C. & Polson, P. (1994) The cognitive walkthrough method: A practitioner's guide. In J. K. Nielson, & R. L. Mack (Eds.) *Usability Inspection Methods*. Wiley, New York, 1994.

[18] Spencer, R. 2000. The streamlined cognitive walkthrough method: Working around social constraints encountered in software development company. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI '00)*, ACM Press, The Hague.

[19] Loewenberg Ball, D., & Forzani, F. M. 2009. The work of teaching and the challenge for teacher education. *Journal of Teacher Education* 60, 5 (2009), 497-511. https://doi.org/10.1177/0022487109348479

[20] Duan, H., Li, J., Fan, S., Lin, Z., Wu, X., & Cai, W. 2021. Metaverse for social good: A university campus prototype. In *Proceedings of the 29th ACM International Conference on Multimedia*, 153-161. https://doi.org/10.1145/3474085.3479238

[21] Dai, C.-P., Ke, F., Dai, Z., & Pachman, M. 2023. Improving teaching practices via virtual reality-supported simulation-based learning: Scenario design and the duration of implementation. *British Journal of Educational Technology* 54, 4 (2023), 836–856. https://doi.org/10.1111/bjet.13296

[22] Veletsianos, G., & Miller, C. 2008. Conversing with pedagogical agents: A phenomenological exploration of interacting with digital entities. *British Journal of Educational Technology* 39, 6 (2008), 969-986. https://doi.org/10.1111/j.1467-8535.2007.00797.x

[23] Rod A Martin, Patricia Puhlik-Doris, Gwen Larsen, Jeanette Gray, and Kelly Weir. 2003. Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality*, 37, (2003), 48–75. https://doi.org/10.1016/S0092-6566(02)00534-2

[24] Ke, F., Yuan, X., Southerland, S. 2021. *Teaching Practices with Multiplayer Mixed Reality Simulations and Virtual Students*. National Science Foundation (Award Abstract # 2110777). Retrieved from https://www.nsf.gov/awardsearch/showAward?AWD\_ID=\$2110777

[25] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. 2019. *Language models are unsupervised multitask learners*. Technical report, OpenAI. https://cdn.openai.com/better-language-models/language\_models\_are\_unsupervised\_multitask\_learners.pdf

[26] Dai, Z., Ke, F., Dai, C.-P., Pachman., M., & Yuan, X. 2021. Role-play in virtual reality: A teaching training design case using OpenSimulator. In G. Akcayir, & C. D. Epp (Eds.), *Designing, Deploying, and Evaluating Virtual and Augmented Reality in Education* (pp.143-163). IGI Global. https://doi.org/10.4018/978-1-7998-5043-4.ch007

[27] Dai, C-P. 2023. Enhancing Learning Achievements and Self-Efficacy for Preservice Teachers Using Model-Based Support in Simulation-Based Learning with Artificial Intelligence-Powered Virtual Agents. Dissertation, Florida State University.

[28] Yin, R. K. 2018. *Case Study Research and Applications: Design and Methods* (6th ed.). Sage, Thousand Oaks, CA.

[29] Ryu, J., & Baylor, A. L. 2005. The Psychometric Structure of Pedagogical Agent Persona. *Technology, Instruction, Cognition & Learning (TICL)* 2, 4 (2005), 291-315.

[30] Glaser, B. G. 1965. The constant comparative method of qualitative analysis. *Social Problems* 12, 4 (April 1965), 436–445. https://doi.org/10.1525/sp.1965.12.4.03a00070.

[31] Dai, C.-P., Ke, F., Dai, Z., West, L., Bhowmik, S., & Yuan, X. 2021. Designing artificial intelligence (AI) in virtual humans for simulation-based training with graduate teaching assistants. In de Vries, E., Hod, Y., & Ahn J. (Eds.), *Proceedings of the 15th International Conference of the Learning Sciences - ICLS 2021*, 1101–1102. Bochum, Germany: International Society of the Learning Sciences. https://repository.isls.org/bitstream/1/7418/1/1101-1102.pdf