

How Persona and Message Type in Conversation Initiation Influence Learners' Experience with Educational Chatbots

Jingchao Fang¹ Qingxiaoyang Zhu² Hao-Chuan Wang²

¹University of Chicago

jcfang@uchicago.edu

²University of California, Davis

{qinzhu, hciwang}@ucdavis.edu

Abstract

Video-based learning is widely adopted among online learners, but it often lacks interactivity and fails to engage learners. Advances in large language models (LLM) show promise in enabling role-playing characters as learning companion chatbots for interactive learning. While much research has focused on LLM-powered chatbots for question-answering or responding, fewer studies have explored their role as proactive conversation initiators and prompt learners to think and reflect. How do different styles of chatbots' conversation initiation influence learners' video-based learning experience? Through controlled empirical studies with learners ($N = 65$), we find that learning companion chatbot's *personas* and *message styles* influence learners' perceptions, reactions, and overall learning experiences. Our findings provide key insights for designing engaging, effective education chatbots.

1 Introduction

Watching recorded educational videos is a common way to learn. While video-based learning excels in accessibility and flexibility, it falls short in interactivity. Passive video-watching hinders learners' deep cognitive engagement with the learning material, increases the likelihood of distraction, and impedes knowledge construction, leading to reduced learning outcomes and lower course completion rates (Schacter and Szpunar, 2015; Alghamdi et al., 2020; Koedinger et al., 2015; Fang et al., 2022; Johnson et al., 2008). Various methods for enhancing the interactivity of online learning have been proposed (Shin et al., 2018; Guo et al., 2014; Fang et al., 2022, 2024; Chi et al., 2025).

Among them, adopting **chatbots as learning companions** has gained increasing attention. Examples include Sofia (Knill et al., 2004) and Ms. Lindquist (Heffernan and Koedinger, 2002) for math education, Sara for programming educa-

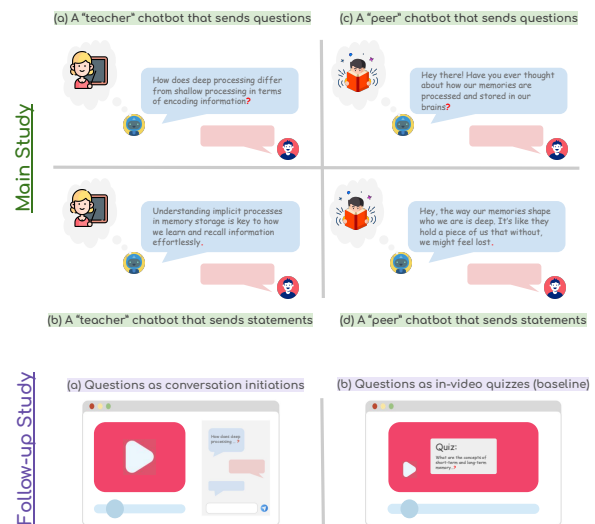


Figure 1: In the main study, we explored how a chatbot's *persona* and *message type* in their conversation initiations affect video-based learners' experiences. In the follow-up study, we compare learning prompts sent as conversation initiations versus learning prompts sent as quiz questions.

tion (Winkler et al., 2020), and Piagetbot for psychology education (Heller and Procter, 2007). Two driving forces of educational chatbot design are the *theory of scaffolding* (Winkler et al., 2020), which argues that learners learn better when being provided with temporal support (Wood et al., 1976; Vygotsky, 1978; Hedegaard, 1990), and the idea of *constructivist learning* (Ruan et al., 2019; Graesser et al., 1999), which emphasizes the importance of reflection and active knowledge construction (Fosnot and Perry, 1996). However, empirical studies testing the utility of educational chatbots have yielded mixed results; one key reason is that the effectiveness of incorporating a chatbot in video-based learning may be influenced by many factors, such as the capability of the underlying algorithmic model (Cai et al., 2024), users' subjective perception (Khadpe et al., 2020), and its interaction mode

and time duration with learners (Winkler et al., 2020), just to name a few.

To bring forward a learner-centric design requires us to understand what characteristics of learning companion chatbots affect learners' experience. Most related works concentrate on chatbots' capability of responding to chats initiated by learners (e.g., question-answering or explaining) (Salminen et al., 2024; Dan et al., 2023; Schmitt et al., 2022; Farah et al., 2022). Recently, there has been growing interest in designing agents that actively extend dialogue or trigger interactions (Hu et al., 2024; Liu et al., 2025). Educational chatbots' role as conversation starters remains underexplored (Park and Kulkarni, 2023). We posit that an important function of chatbots, as learning companions, is to **proactively prompt interactions** rather than wait to be called upon. To advance our understanding in this emerging area, we investigate how different ways that a chatbot's **conversation initiation** affects learners' reactions and perceptions.¹ This scope might appear small but non-trivial—though brief, conversation initiation plays a critical role in the overall interaction as it sets the tone for everything that follows, for example, by establishing relational dynamics and framing the expectation (Schegloff, 1968).

We focus on two factors that shape a learning companion chatbot's initiation message: the **persona** of the chatbot, (i.e., the distinct characteristic or role that the chatbot embodies) and the **message type** (i.e., the communicative function or format of the message). Our decision to focus on these factors stems from the essential roles they play in education contexts: (1) learners are typically aware of the identity of the individuals they interact with—most often teachers/experts or fellow learners/peers; (2) conversations during teaching and learning commonly involve asking questions or making statements. Existing learning companion chatbots often follow these patterns, disclosing their roles and engaging in conversations through questions or statements. A chatbot's persona and message type largely determine its style of engagement and are among the most salient cues in learner-chatbot interaction, offering us a practical lens to explore the utility and strategic design of educational chatbots.

We developed and used EduPrompt, a conversa-

tion user interface, as a *technology probe* (Hutchinson et al., 2003) to conduct a 2×2 controlled experiment, investigating the effects of a chatbot's persona (a *teacher/peer*) and message type (sending *questions/statements*) when initiating new conversations. In a follow-up study, we compared the chatbot's conversation initiations with in-video quiz questions (baseline condition), as embedding quizzes is a widely adopted approach to enhance learning interactivity in online education platforms (e.g., Coursera).² Figure 1 summarizes the experiments.

Our **contribution** is two-fold: (1) We provide empirical results that reveal the utility of educational chatbots and how their persona and message type influence users' learning experience. For example, we show that questions framed as conversation initiations resulted in learners' lower cognitive load compared to quiz-style questions, and a chatbot that role-plays a peer may be more favorable than a chatbot that role-plays a teacher in some circumstances; (2) We contribute a learner-centered design perspective for enhancing chatbots' utility and offer concrete suggestions for designing learning companion chatbots to foster engagement in video-based learning.

2 Background

2.1 Learning Prompts and Chatbots for Online Video-Based Learning

Learners learn more effectively when they are prompted to explain the learning material (Chi et al., 1994; Aleven et al., 2004; Williams and Lombrozo, 2010) (see additional details in Appendix A). In-video learning prompts are adopted in some online learning platforms (e.g., Coursera) in the form of multiple-choice quiz questions. However, prompting learners with quiz questions may also cause learners to ignore details and lead to anxiety (Williams and Lombrozo, 2010; Shin et al., 2018). Engaging in conversations with others can offer benefits similar to learning prompts while enhancing interactivity, effectiveness, and joy in learners' experience (Cortright et al., 2005; Ibrahim and Al-Shara, 2007), but it is constrained by the availability of a chatting partner. Educational chatbots have gained attention for their potential to serve as always-available learning companion to be interacted with.

¹In this paper, we define conversation initiation as the first message sent by one party and (optionally) the first response sent by the other.

²<https://www.coursera.org/>. Accessed in 2026.

Most existing educational chatbots take an authoritative or guiding role (e.g., a teacher or a teaching assistant) and deliver pedagogical materials and feedback, such as AutoTutor that teaches Newtonian qualitative physics and computer literacy (Graesser et al., 2004), MathBot that explains math concepts and offers feedback to students (Grossman et al., 2019), Quizbot that helps learners memorize factual knowledge (Ruan et al., 2019), and Jill Watson deployed as a community-facing virtual teaching assistant (Wang et al., 2021). Several studies designed virtual co-learners that go beyond knowledge-based assistance by providing scaffolding through social support. In a VR classroom that Liao et al. (2019) designed, learners can watch lecture videos with virtual classmates embodied with historical learners’ messages (time-anchored comments). More recently, Wang et al. (2025) introduced generative co-learners that engage in multimodal interactions with learners to enhance their social awareness during learning. Little attention has been paid to educational chatbots’ role in proactively initiating interactions. Initial efforts include building models that lead knowledge-grounded conversations (Wu et al., 2019), which focus on improving the technological capability of interactive systems. In our study, we use LLM to generate context-appropriate conversation initiation and focus on understanding the utility of applying chatbots’ proactive conversation initiation in educational contexts.

2.2 Effects of Different Messaging Styles

Educational chatbots designed and studied in prior works exhibit considerable diversity, sending messages that vary in choice of language (formal or informal), content focus (factual, reflective, motivational), and abstraction levels (simple or complex) (Tannen, 2005). These variations led to inconsistent effects observed in learning outcomes and learners’ experiences when using chatbot-based learning support systems (Ruan et al., 2019; Schroeder and Adesope, 2014). To disentangle the twisted factors that exist in the complex learner-chatbot interaction dynamics, in this paper, we investigate the different effects of a chatbot’s *persona* and *messages type* in a strictly controlled experiment.

The persona of a chatbot, manifested through personality traits, tone, and so on, may shape how users perceive and interact with the chatbot (Khadpe et al., 2020), making it a critical area

of study. This phenomenon, known as the *persona effect*, has been studied extensively in educational contexts (Lester et al., 1997). Through assessing middle school students’ perception of animated pedagogical chatbots’ affective characteristics, Lester et al. (1997) found that “the presence of a lifelike character in an interactive learning environment—even one that is not expressive—can have a strong positive effect on student’s perception of their learning experience”. Simulating personas and adapting to diverse application scenarios is easier than ever with recent advances in LLM-powered role-playing agents (Tseng et al., 2024; Tang et al., 2024; Fu et al., 2023; Zhong et al., 2020).

The type of messages sent by a chatbot may also impact how learners interact with it. In education scenarios, most learning prompts are presented as questions, which are designed to encourage learners’ self-explanation, assess their understanding, or identify knowledge gaps (Williams and Lombrozo, 2010; Ruan et al., 2019). However, when learners interact with teachers or peers in real life, they often also receive statements. For example, learners may receive instructions from teachers or learning summaries and thoughts shared by peers (Fang et al., 2022). Questions and statements can serve different functions and induce different reactions in dialogues (Searle, 1969). This study investigates how the message type of a chatbot’s conversation initiation—questions versus statements—affects learners’ experiences.

3 Experiment Tool: EduPrompt

We prototyped EduPrompt as an experiment tool to manipulate participants’ study conditions. EduPrompt is implemented as a web application that embeds an educational video and a chatbox side by side (Appendix B Figure 3). A chatbot named Alex sends conversation initiation messages in the chatbox,³ prompting users to engage with the learning video segment by segment.

EduPrompt follows a pipeline to automatically generate the prompting messages. At the beginning of each learning session, the chatbot sends a welcome and self-introduction message to the learner: “Hello, I am Alex, a teacher (or your peer learner) of this class. Welcome to EduPrompt! Please start the video.” Backend retrieves metadata and the full transcript of

³We intentionally picked a gender-neutral name to avoid introducing a confounding variable.

the video through the YouTube Transcript/Subtitle API.⁴ We segment the video based on the ending time of video chapters.

To minimize user distraction, EduPrompt only sends messages at the end of video segments; messages are synthesized by GPT-3.5 following a system prompt template (Appendix C). The two factors we study, persona and message type, are embedded into the template to ensure the generated messages match Alex's assigned message style. Because this study focuses on conversation initiation, each message generation is solely based on the learning material of the current video segment (represented by the retrieved transcript); additional contextual information, such as the chat history during previous video segments, is not considered. EduPrompt automatically pauses the video when the message is sent out. If the learner responds, the chatbot will reply with a uniform message ("Great, let's move on. Please continue to watch the video.") Otherwise, learners can ignore the message by directly restarting the video without responding. We intentionally disabled back-and-forth conversations between the chatbot and the learner, because isolating the initiation stage of chatbot-learner conversations from the rest of the conversation is required for a controlled study; multi-round conversations will challenge the control by introducing confounding variables after the initial message is sent.

4 Method

We aim to answer two research questions: **RQ1.** How does a learning companion chatbot that periodically initiates conversations with learners influence their video-based learning experience? **RQ2.** How does a chatbot's messaging style (persona and message type) affect video-based learners' perceptions and reactions? We rolled out the IRB-approved study in two rounds. Round 1: the main study investigates the effects of chatbots' persona and message type; Round 2: the follow-up study compares questions as conversation initiations versus questions as quizzes (baseline condition).

4.1 Main Study

Factors. We consider two independent variables: the chatbot's persona and message type. Regarding persona, the chatbot role-plays either a teacher or a

peer of the educational video, which decides how it introduces itself at the beginning of the learning session and how it frames initiation messages to kick off conversations after each video segment. Regarding message type, the chatbot's messages (except the initial welcome message) are either all questions or all statements. We selected two educational videos as learning materials in EduPrompt.⁵ Both videos are about 10 minutes long and are delivered in a popular science style, making them suitable for a general audience and requiring no prior knowledge to understand. Each participant watches one of the two selected videos during their study session and receives messages from a learning companion chatbot. We treat the video content as a random factor.

Message generation and manipulation check.

Due to ethical considerations for human-subject research and the goal of strict experiment control, we pre-generated all Alex's messages using the prompt template (Appendix C) and stored them in a JSON file before study deployment. During each participant's study session, the messages were retrieved at the end of a video segment and posted in the EduPrompt chatbox. Researchers conducted a manipulation check to ensure that (1) all generated messages were appropriate for human-facing studies; (2) the LLM followed instructions in the system prompt and generated significantly distinct messages when instructed to adapt message style based on personas and message types. Appendix D includes additional details about the manipulation check and examples of messages that participants received (Table 1).

Study design. We conducted a 2×2 between-subjects experiment and collected data from 43 participants. Participants were randomly assigned to one of the four conditions: 11 participants received questions from Alex as a peer (labeled as P1-P11), 11 participants received questions from Alex as a teacher (P12-P22), 11 participants received statements from Alex as a peer (P23-P33), and 10 participants received statements from Alex as a teacher (P34-P43), yielding a relatively balanced sample size for each condition. The study takes approximately 35 minutes, which includes reading

⁴<https://pypi.org/project/youtube-transcript-api/> Accessed in 2026.

⁵"How We Make Memories" (<https://www.youtube.com/watch?v=bSyedIx-C48>), accessed in 2026; "What Does the Atmosphere Do" (<https://www.youtube.com/watch?v=11ZI9aqrFA&list=PL8dPuuaLjXtO85S124rSiVQ93q7vCntNF&index=7>), accessed in 2026.

a consent form and a study instruction, learning with a video using EduPrompt, and filling out a post-task survey. Details about the recruitment can be found in Appendix E.

Measurements. The EduPrompt web application logs all participants' behavioral data (e.g., start and finish time, chat history). In the post-task survey, participants rate their emotional engagement, cognitive engagement, self-efficacy in learning, perceived distraction, cognitive load, and perceived rapport with the chatbot using 7-point Likert scale items, and they rate the chatbot's anthropomorphism, intelligence, and likeability using 5-point Likert-scale items. The survey also asks for participants' demographic information. Participants elaborate on their feedback by responding to a series of open-ended questions. All survey items and sources are listed in Appendix F.

4.2 Follow-Up Study

To encourage learners' reflection and enhance cognitive engagement, the conventional approach adopted by many platforms (e.g., Coursera and edX)⁶ is to embed quiz questions in videos. How do framing questions as conversation initiations, sent by a role-playing chatbot, differ from presenting them as quizzes? We conducted a follow-up study as the baseline condition with an additional group of 22 participants (labeled as P44-P65). The recruitment method, study procedure, and learning material are similar to those of the main study. There were only three changes: (1) we called GPT-3.5 API using adjusted system prompts to generate questions as quizzes, removing any persona or humanlike characteristics; (2) the chatbox was removed from the front-end interface, and each question was rendered in an in-video quiz box during the pause at the end of the corresponding video segment; (3) the survey items solely about perceptions of the learning companion chatbot (including perceived rapport, Alex's anthropomorphism, intelligence, and likeability, and the open-ended feedback) were removed, while other items that mentioned Alex were adapted to match the study context.

5 Quantitative Analysis and Result

Our analysis primarily relies on self-reported measures because this study focuses on learners' *sub-*

jective learning experiences rather than their *objective learning outcomes*. Self-reports provide direct insight into the learners' subjective perceptions. To triangulate these findings, we incorporate high-level behavioral metrics (e.g., word counts of participants' messages). We deferred granular content analysis of participants' chat history due to the high variance inherent in linguistic data, which could obscure meaningful patterns when the sample size is not large.

Given the relatively small sample size, we report both statistical significance (p -value) and effect size (η_p^2) for all comparisons. This follows established best practices, as p -values alone are highly sensitive to sample size and effect size is better suited to provide an indication of the magnitude of the observed effects (Bakker and Wicherts, 2011; Berben et al., 2012).

Comparing chatbots' conversation initiations with varying Persona and message type.

We conducted a two-way ANOVA to compare the effects caused by the learning companion chatbot's different personas and message types when initiating conversations. We found that the chatbot's persona may shape learners' overall experience. Learners perceived a significantly stronger rapport with the chatbot when it roleplays a peer to initiate conversations, compared to when it roleplays a teacher ($F[1, 39] = 5.04, p < .05, \eta_p^2 = .11$; chatbot as a teacher: $M = 3.7, SD = 1.78$; chatbot as a peer: $M = 4.9, SD = 1.59$). Learners also reported the chatbot to have a higher level of anthropomorphism when it speaks in a peer's tone instead of a teacher's tone ($F[1, 39] = 6.36, p < .05, \eta_p^2 = .14$; chatbot as a teacher: $M = 2.44, SD = 1.13$; chatbot as a peer: $M = 3.43, SD = 1.34$). In addition, they also rated the chatbot as more intelligent when it roleplays a peer compared to when it roleplays a teacher, with borderline significance observed ($F[1, 39] = 3.92, p = .05, \eta_p^2 = .09$; chatbot as a teacher: $M = 3.60, SD = 1.00$; chatbot as a peer: $M = 4.20, SD = 0.92$). No significant difference was observed regarding the chatbot's likeability. We did not find evidence that chatbot's persona and message type influence learners' self-rated emotional engagement, cognitive engagement, self-efficacy of learning, perceived distraction, and cognitive load in the learning task (all $p > .05$).

Learners' responses to conversation initiations also suggest that message type could affect their

⁶<https://www.edx.org/> Accessed in 2026.

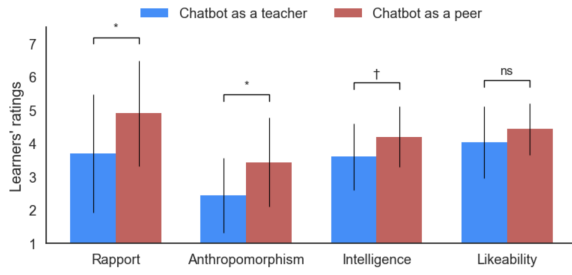


Figure 2: Learners' ratings on the chatbot's perceived rapport, anthropomorphism, intelligence, and likeability. ns: not significant; †: borderline significant with $p = 0.05$; *: significant with $p < 0.05$. Error bars represent standard deviations.

reactions to the learning companion chatbots. One of the videos has 10 segments, and the other has 11 segments; thus, the chatbot sends 9 and 10 messages (in addition to the initial welcome message) during the video-playing accordingly. Participants send 9.26 messages on average ($min = 3, max = 10, med = 10, SD = 3.18$), and the average length of messages is 21.18 words or 121.3 characters. The chatbot's persona and message type do not have a main effect on how likely learners are to respond, according to the number of messages participants send during video-watching. However, participants who received questions sent significantly longer messages than participants who received statements (in number of words: $F[1, 39] = 9.30, p < .01, \eta_p^2 = .19$; when messages are questions: $M = 29.17, SD = 21.97$; when messages are statements: $M = 12.80, SD = 9.94$; in number of characters: $F[1, 39] = 9.86, p < .01, \eta_p^2 = .20$; when messages are questions: $M = 168.78, SD = 126.05$; when messages are statements: $M = 71.58, SD = 56.91$).

Comparing questions as conversation initiation with questions as quizzes. We compared quantitative self-reported data from the main study participants assigned to receive questions from Alex (either as a teacher or a peer) with data from all follow-up study participants. Statistics showed that learners who received questions presented as in-video quizzes reported significantly higher cognitive load than those who received questions as conversation initiation messages ($F[1, 42] = 5.09, p < .05, \eta_p^2 = .11$; questions as initiation messages: $M = 2.79, SD = 1.12$; questions as quizzes: $M = 3.67, SD = 1.45$). Although learners' actual question response rate did not differ (according to the behavioral log), learners who

answered quiz questions self-rated significantly higher frequency of engagement than learners who replied to message questions ($F[1, 42] = 4.17, p < .05, \eta_p^2 = .09$; questions as initiation messages: $M = 5.0, SD = 1.8$; questions as quizzes: $M = 6.0, SD = 1.25$), which aligns with the observed difference in cognitive load. ANOVA did not show differences in learners' perceived engagement, number of responses to initiation messages, and perceived helpfulness of the tool ($p > .05$).

6 Qualitative Analysis and Learners' Feedback

We conducted a thematic analysis on participants' responses to the open-ended survey questions (Braun and Clarke, 2012). The findings revealed users' elaborative feedback on how an educational chatbot that periodically initiates conversations may help or impede their learning.

A chatbot that initiates conversations could improve learners' experience. Most participants enjoyed learning with the learning companion chatbot. They repeatedly mentioned several advantages of interacting with the chatbot:

Sustain learners' attention. Many participants stated that the chatbot's messages helped them to stay focused on the video. P5 stated, "*The interactions kept me focused on the video and made me think about what I was learning. [...] the frequent breaks from Alex allowed my brain time to process what I had learned and to think more about it.*" The chatbot helped them recognize and focus on the important information. P21 mentioned that, "*it allowed me to try and focus better in the important information so I can answer the questions.*" Some participants attributed this positive effect to the automatic pauses that broke learning videos into smaller pieces. For example, P22 said, "*[It] helps me learn piece by piece instead of all at once.*"

Prompt learners to think and reflect, and help with knowledge retention. Participants mentioned that the chatbot's messages triggered in-depth thinking and reflection on the video content. P5 mentioned, "*it forced me to think more about what I was watching. Alex would be a very helpful tool for learning retention.*" P12 also described EduPrompt as "*a tool that was trying to have me recollect material I learned*".

Overall, participants commended EduPrompt for **making the learning experience interactive and fun**. P38 stated, "*It felt like I had someone else*

there with me learning and coaching me through it". P32 also mentioned, "It helps me to interact and improve the learning process better than my previous learning method."

Negative side effects. Participants also mentioned drawbacks of learning with a proactive chatbot. Two themes emerged from their responses:

Frequent interruptions. The videos we selected for the user studies include many segments; thus, the videos were frequently automatically paused. Learners who showed negative attitudes toward the chatbot mostly complained about the frequent pauses. P33 mentioned *"Alex messaged too often [...] it kind of broke up the momentum."*

Chatbot's unaltered replies. Several participants were disappointed by how the chatbot replied to their messages. P35 mentioned, *"[...] the chatbot was not very dynamic and did not answer queries in any satisfying way."* This is a side effect of our effort to restrict the conversation to a single round for strict experiment control. This should not be a problem in real-world scenarios of integrating a fully functional chatbot into video-based learning.

Considerations when learning with longer or more educational videos. Due to practical limitations in conducting a user study with online participants, we only included one short video in each study session. Many educational videos are longer, and learners sometimes learn through a series of videos rather than a single one. We captured learners' thoughts about deploying educational chatbots in longer or more video-based learning sessions.

Potentially leading to changes in messaging behaviors. Some learners expected that they would change their messaging behavior when the learning tasks are heavier. P18 mentioned that *"I would probably answer the questions a lot less often as answering the questions thoughtfully added a lot of additional time to watching the video."* A few participants suggested that the pause frequency should be reduced for longer videos.

Enhanced interactivity may become more helpful when learning with longer or more videos. Despite several participants worrying about the increased workload, others think that the enhanced interactivity becomes more pronounced when learners watch longer or more videos. P42 suggested that *"When watching long videos, this tool makes it less stressful and boring because of the chat feature [...]"*. P6 mentioned that, when learning through a series of videos, *"I would lean toward more willing*

[to learn with EduPrompt] because I appreciated the interactive experience here and how it was broken down into key segments ..."

7 Discussion

Online education carries the promise of making knowledge accessible to all, but its reduced interactivity compromises learners' experience, preventing it from truly closing the gap with traditional face-to-face education. Deploying LLM-powered proactive learning companion chatbots may boost interactivity in online video-based learning, but how should we design such chatbots? In today's world where LLMs are predominantly used in education to offload critical thinking, leading to cognitive debt and diminished learning gains (Kosmyna et al., 2025), this investigation is especially timely as it demonstrates an alternative use case for LLMs which introduces "positive friction" and "desirable difficulty" in the learning process (Bjork et al., 2011; Collins et al., 2024), rather than making learning easier and smoother.

7.1 Summary of Findings

A learning companion chatbot that proactively prompts interactions could enrich video-based learning. (RQ1). Empirical results showed that it interrupted learners' mindless video-watching by breaking videos into smaller, more digestible pieces. Context-aware conversation initiations sustained learners' attention, helped learners identify important content, and encouraged learners to deeply engage with the videos through reflective thinking. While our study focused on measuring learning experiences rather than learning outcomes, we infer that learning with a learning companion chatbot could be pedagogically beneficial according to the ICAP framework, which indicates that interactive learning leads to better knowledge retention and deeper understanding than passive learning (Chi and Wylie, 2014).

We highlight the utility of presenting learning prompts as conversation initiations sent by a role-playing chatbot. Compared to embedding quiz questions into videos, framing questions as conversation initiations results in significantly lower cognitive load without reducing learners' engagement. We expect that substituting quizzes with conversational questions may be especially beneficial for learners who are less familiar with the learning material, as research indicates that they

perceive a higher cognitive load and anxiety than more knowledgeable learners when taking in-video quizzes (Shin et al., 2018).

The persona and message type of a learning companion chatbot significantly influence learners' experience. (RQ2). The persona of the learning companion chatbot and the type of message it sends affect how people perceive and interact with it. Learners perceived the chatbot to have higher levels of anthropomorphism and intelligence, two highly desired characteristics in human-chatbot interaction (Wang et al., 2021), when it role-plays a peer instead of a teacher. Learners also rated a “peer” chatbot to be more capable of building rapport than a “teacher” chatbot. This might be caused by people’s higher expectations of a teacher’s role. Some learners may expect a “teacher” chatbot to be more authoritative or provide feedback with learners’ messages (e.g., whether an answer is right or wrong). Learners’ varying mental models and expectations of personas shape their differing perceptions of the chatbot they interacted with.

Learners constructed longer messages when responding to questions than statements sent by the chatbot. A number of learners also mentioned that the messages they jotted down could be used as learning notes for future review. In this sense, a chatbot that asks questions may be more helpful for learners to engage deeply with the learning materials and construct more comprehensive and reflective insights that support long-term understanding and retention.

We discuss additional considerations for deploying educational chatbots to support diverse learning scenarios in Appendix G.

7.2 Design Implications

To focus on the effects of the chatbot’s embodied persona and message type on conversation initiations, we disabled the chatbot from responding to learners. However, the end goal of this exploration is to probe the design space for full-fledged chatbots that engage with multi-turn conversations and serve as learning companions when people passively watch educational videos. Based on our empirical results, we discuss considerations when designing fully functional educational chatbots.

Choosing the right persona. The initiation of conversations shapes users’ first impression of a

chatbot and significantly influences how users perceive it. A large body of work has highlighted the crucial role that a persona or metaphor plays in human-chatbot interaction (Khadpe et al., 2020; Liu et al., 2022; Park et al., 2024). Our study further confirmed that *strategically selecting a persona for the chatbot to roleplay is essential for fostering positive learning experiences*. The persona effect is significant even when it is solely embedded in the conversation initiations. Most educational chatbot products today role-play a “teacher” or a “tutor”; however, not all chatbots are expected or equipped to give definitive responses or authoritative guidance to the learners. In such cases, we suggest that chatbot designers avoid using an “expert” persona (e.g., a teacher, an experienced professional in the field). Instead, using a non-expert persona (e.g., a peer or classmate who is also learning) may align better with learners’ expectations.

Deciding on the right message type. We suggest chatbot designers taking a goal-oriented approach when deciding what type of messages their chatbots send. For learning companion chatbots, designers should *manipulate the chatbot’s message type depending on learners’ expected outcome of interacting with the chatbot*. For example, if learners are supposed to construct concrete, detailed messages to externalize their knowledge as study notes, a question-asking chatbot could facilitate them to achieve this outcome. In other scenarios where the chatbot merely serves as a light intervention, and learners are not expected to spend much time conversationally engaging with the chat, a chatbot that periodically delivers statement messages could effectively meet this objective.

8 Conclusion

We explore educational chatbots’ effects on video-based learners’ experience when they act as conversation initiators. Through controlled experiments, we find that the persona of a learning companion chatbot and the type of messages it sends can lead to learners’ different reactions and perceptions. We also find that questions framed as conversation initiations resulted in significantly lower cognitive load compared to those presented as quizzes, validating the benefit of integrating learning companion chatbots to promote relaxed engagement. We provide key insights for designing educational chatbots that enrich video-watching experiences by prompting meaningful engagement.

Limitations

There are several limitations of our study. Due to the challenges inherent in conducting any experiments with online participants, we intentionally designed short user studies, each featuring only one short video per learning session. While we asked learners' opinions about using EduPrompt in scenarios involving longer or multiple videos, the insights they provided remain speculative and are not grounded in direct observation.

Our current study focuses on understanding learners' learning experience and perceptions rather than their learning outcomes. Notably, significant differences in knowledge gain are rarely observable within a single, short-duration learning session. Future work involving longer, multiple learning sessions will be necessary to examine whether the experiential advantages we observed in our study could eventually translate into measurable improvements in learning outcomes.

We acknowledge the limitations posed by our relatively small sample size and have taken appropriate steps to ensure a realistic interpretation of our quantitative results. This included reporting effect sizes alongside statistical significance to quantify the magnitude of observed effects and assess their practical significance independent of sample size. To triangulate these quantitative findings, we incorporated detailed qualitative analysis to capture nuanced participants' perceptions. Given that conversational data can exhibit high inherent variance within the small sample, we omitted granular textual analysis of participants' chat history. Future research can replicate and expand the study to allow for a more detailed exploration of how a chatbot's persona and message type in conversation initiations may affect participants' dialogue patterns.

We used GPT-3.5 to generate the learning companion chatbot's conversation initiation messages. Since the task only involves basic role-playing and message synthesis based on introductory-level educational contexts, we do not anticipate that using a more recent or advanced model would yield substantially different results. That being said, in education scenarios where initiating a conversation requires more complex reasoning or the ability to process longer context, comparing outputs from different models may be worthwhile.

Ethical and Societal Implications

Ensuring equitable access to high-quality education remains a critical societal challenge. Online education, to some extent, has democratized learning opportunities by lifting geographical and financial barriers of traditional in-classroom education. However, the effectiveness of online education is constrained by its limited interactivity (e.g., passive video-watching), leaving learners disengaged, and preventing online education from realizing its full potential. NLP can drive positive societal impact by fostering scalable, interactive online education, thereby advancing educational equality.

This work aligns with the mission of “**NLP for Positive Impact**” by addressing a fundamental design challenge in this domain—how to build proactive learning companion chatbots that meaningfully enhance online learners' video-based learning experience. By adopting a learner-centric perspective to examine how chatbot persona and message type during conversation initiation affect learners' experience, our work provides actionable insights for leveraging NLP to foster engaging and enjoyable online video-based education. This work is especially crucial at a time when LLMs are widely adopted by learners to offload their cognitive tasks, rather than facilitate them to think and reflect more critically (Kosmyna et al., 2025).

Our study design and data collection procedures strictly follow ethical best practices for research involving human participants. We obtained approval from IRB prior to data collection, and all study participants provided informed consent before the start of their study session.

We opted to use pre-generated conversation initiation messages and quiz questions as learning prompts in our user studies, even though technically, they can be automatically generated in real-time using the EduPrompt pipeline. This allows us to perform a manipulation check and ensure the generated texts are unharmed and appropriate for being presented in human-facing experiments. This approach did not sacrifice the ecological validity of the study (compared to real-time message/quiz question generation) as we did not alter the generated texts when integrating them into the interface.

We suggest chatbot designers perform a risk assessment before deploying a fully functional chatbot that engages with learners in multi-turn conversations in real-time. Despite significant advances in LLMs, their generated content may still be prob-

lematic. For example, LLM-powered chatbots may hallucinate and generate incorrect or irrelevant responses to a learner's query on related knowledge (McKenna et al., 2023), send messages that include biased content or incoherent or nonsensical text that confuses learners (Fang et al., 2023; Wu et al., 2023). The LLM powering the chatbots may occasionally fail to follow system prompts (Li et al., 2023), leading to inconsistent role-play as a learning companion. While these problems can be mitigated by strategies such as fine-tuning (Peng et al., 2023), to date, there is no practical method for avoiding them entirely. Chatbot designers should carefully assess the risks before deploying LLM-powered learning companion systems at scale and apply strategies to mitigate them.

References

- Vincent Aleven, Amy Ogan, Octav Popescu, Cristen Torrey, and Kenneth Koedinger. 2004. [Evaluating the effectiveness of a tutorial dialogue system for self-explanation](#). In *Intelligent Tutoring Systems: 7th International Conference, ITS 2004, Maceió, Alagoas, Brazil, August 30-September 3, 2004. Proceedings 7*, pages 443–454. Springer.
- Ahlan Alghamdi, Aryn C Karpinski, Andrew Lepp, and Jacob Barkley. 2020. [Online and face-to-face classroom multitasking and academic performance: Moderated mediation with self-efficacy for self-regulated learning and gender](#). *Computers in Human Behavior*, 102:214–222.
- Marjan Bakker and Jelte M Wicherts. 2011. The (mis) reporting of statistical results in psychology journals. *Behavior research methods*, 43(3):666–678.
- Lut Berben, Susan M Sereika, and Sandra Engberg. 2012. Effect size estimation: methods and examples. *International journal of nursing studies*, 49(8):1039–1047.
- Elizabeth L Bjork, Robert A Bjork, et al. 2011. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, 2(59-68):56–64.
- Virginia Braun and Victoria Clarke. 2012. *Thematic analysis*. American Psychological Association.
- Gerald F Burch, Nathan A Heller, Jana J Burch, Rusty Freed, and Steve A Steed. 2015. [Student engagement: Developing a conceptual framework and survey instrument](#). *Journal of Education for Business*, 90(4):224–229.
- Zhenyao Cai, Seehee Park, Nia Nixon, and Shayan Doroudi. 2024. [Advancing knowledge together: Integrating large language model-based conversational ai in small group collaborative learning](#). In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–9.
- Micheline T. H. Chi and Ruth Wylie. 2014. [The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes](#). *Educational Psychologist*, 49(4):219–243.
- Micheline TH Chi, Nicholas De Leeuw, Mei-Hung Chiu, and Christian LaVancher. 1994. [Eliciting self-explanations improves understanding](#). *Cognitive science*, 18(3):439–477.
- Peggy Chi, Senpo Hu, Lei Shi, Tanya Kraljic, Justin Secor, Tao Dong, Irfan Essa, and Mike Cleron. 2025. Watchwithme: Llm-based interactive guided watching of review videos. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces*, pages 1–15.
- Katherine M Collins, Valerie Chen, Ilia Sucholutsky, Hannah Rose Kirk, Malak Sadek, Holli Sargeant, Ameet Talwalkar, Adrian Weller, and Umang Bhatt. 2024. Modulating language model experiences through frictions. *arXiv preprint arXiv:2407.12804*.
- Ronald N Cortright, Heidi L Collins, and Stephen E DiCarlo. 2005. [Peer instruction enhanced meaningful learning: ability to solve novel problems](#). *Advances in physiology education*, 29(2):107–111.
- Yuhao Dan, Zhikai Lei, Yiyang Gu, Yong Li, Jianghao Yin, Jiaju Lin, Linhao Ye, Zhiyan Tie, Yougen Zhou, Yilei Wang, et al. 2023. Educhat: A large-scale language model-based chatbot system for intelligent education. *arXiv preprint arXiv:2308.02773*.
- Jingchao Fang, Jeongeon Park, Juho Kim, and Hao-Chuan Wang. 2024. [EduLive: Re-creating cues for instructor-learners interaction in educational live streams with learners' transcript-based annotations](#). *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2):1–33.
- Jingchao Fang, Yanhao Wang, Chi-Lan Yang, Ching Liu, and Hao-Chuan Wang. 2022. [Understanding the effects of structured note-taking systems for video-based learners in individual and social learning contexts](#). *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP):1–21.
- Xiao Fang, Shangkun Che, Minjia Mao, Hongzhe Zhang, Ming Zhao, and Xiaohang Zhao. 2023. [Bias of ai-generated content: An examination of news produced by large language models](#). *Preprint*, arXiv:2309.09825.
- Juan Carlos Farah, Basile Spaenlehauer, Sandy Ingram, and Denis Gillet. 2022. A blueprint for integrating task-oriented conversational agents in education. In *Proceedings of the 4th conference on conversational user interfaces*, pages 1–8.

- Catherine Twomey Fosnot and Randall Stewart Perry. 1996. Constructivism: A psychological theory of learning. *Constructivism: Theory, perspectives, and practice*, 2(1):8–33.
- Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. 2023. Improving language model negotiation with self-play and in-context learning from ai feedback. *arXiv preprint arXiv:2305.10142*.
- Arthur C Graesser, Shulan Lu, George Tanner Jackson, Heather Hite Mitchell, Mathew Ventura, Andrew Olney, and Max M Louwerse. 2004. *Autotutor: A tutor with dialogue in natural language*. *Behavior Research Methods, Instruments, & Computers*, 36:180–192.
- Arthur C Graesser, Katja Wiemer-Hastings, Peter Wiemer-Hastings, Roger Kreuz, Tutoring Research Group, et al. 1999. *Autotutor: A simulation of a human tutor*. *Cognitive Systems Research*, 1(1):35–51.
- Joshua Grossman, Zhiyuan Lin, Hao Sheng, Johnny Tian-Zheng Wei, Joseph J Williams, and Sharad Goel. 2019. Mathbot: Transforming online resources for learning math into conversational interactions. *AAAI 2019 Story-Enabled Intelligence*.
- Philip J. Guo, Juho Kim, and Rob Rubin. 2014. *How video production affects student engagement: An empirical study of mooc videos*. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, page 41–50, New York, NY, USA. Association for Computing Machinery.
- Sandra G Hart and Lowell E Staveland. 1988. *Development of nasa-tlx (task load index): Results of empirical and theoretical research*. In *Advances in psychology*, volume 52, pages 139–183. Elsevier.
- Mariane Hedegaard. 1990. *The zone of proximal development as basis for instruction*, page 349–371. Cambridge University Press.
- Neil T Heffernan and Kenneth R Koedinger. 2002. *An intelligent tutoring system incorporating a model of an experienced human tutor*. In *International Conference on Intelligent Tutoring Systems*, pages 596–608. Springer.
- Bob Heller and Mike Procter. 2007. Conversational agents and learning outcomes: An experimental investigation. In *EdMedia+ Innovate Learning*, pages 945–950. Association for the Advancement of Computing in Education (AACE).
- Jiaxiong Hu, Jingya Guo, Ningjing Tang, Xiaojuan Ma, Yuan Yao, Changyuan Yang, and Yingqing Xu. 2024. *Designing the conversational agent: Asking follow-up questions for information elicitation*. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1):1–30.
- Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, et al. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 17–24.
- Mohamed Ibrahim and Osama Al-Shara. 2007. *Impact of interactive learning on knowledge retention*. In *Human Interface and the Management of Information. Interacting in Information Environments: Symposium on Human Interface 2007, Held as Part of HCI International 2007, Beijing, China, July 22-27, 2007, Proceedings, Part II*, pages 347–355. Springer.
- Richard D Johnson, Steven Hornik, and Eduardo Salas. 2008. *An empirical examination of factors contributing to the creation of successful e-learning environments*. *International Journal of Human-computer studies*, 66(5):356–369.
- Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T Hancock, and Michael S Bernstein. 2020. *Conceptual metaphors impact perceptions of human-ai collaboration*. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–26.
- Oliver Knill, Johnny Carlsson, Andrew Chi, and Mark Lezama. 2004. An artificial intelligence experiment in college math education. *Preprint available at <http://www.math.harvard.edu/~knill/preprints/sofia.pdf>*.
- Kenneth R Koedinger, Jihee Kim, Julianna Zhuxin Jia, Elizabeth A McLaughlin, and Norman L Bier. 2015. *Learning is not a spectator sport: Doing is better than watching for learning from a mooc*. In *Proceedings of the second (2015) ACM conference on learning@ scale*, pages 111–120, New York, NY, USA. Association for Computing Machinery.
- Nataliya Kosmyrna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes. 2025. Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, 4.
- James C. Lester, Sharolyn A. Converse, Susan E. Kahler, S. Todd Barlow, Brian A. Stone, and Ravinder S. Bhogal. 1997. *The persona effect: affective impact of animated pedagogical agents*. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, page 359–366, New York, NY, USA. Association for Computing Machinery.
- Shiyang Li, Jun Yan, Hai Wang, Zheng Tang, Xiang Ren, Vijay Srinivasan, and Hongxia Jin. 2023. *Instruction-following evaluation through verbalizer manipulation*. *Preprint*, arXiv:2307.10558.
- Meng-Yun Liao, Ching-Ying Sung, Hao-Chuan Wang, and Wen-Chieh Lin. 2019. *Virtual classmates: Embodying historical learners' messages as learning companions in a vr classroom through comment mapping*. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*, pages 163–171. IEEE.

- Junfeng Liu, Christopher Symons, and Ranga Raju Vatsavai. 2022. [Persona-based conversational ai: State of the art and challenges](#). In *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 993–1001. IEEE.
- Xingyu Bruce Liu, Shitao Fang, Weiyan Shi, Chien-Sheng Wu, Takeo Igarashi, and Xiang'Anthony' Chen. 2025. Proactive conversational agents with inner thoughts. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–19.
- Nick McKenna, Tianyi Li, Liang Cheng, Mohammad Javad Hosseini, Mark Johnson, and Mark Steedman. 2023. [Sources of hallucination by large language models on inference tasks](#). *arXiv preprint arXiv:2305.14552*.
- Gain Park, Jiyun Chung, and Seyoung Lee. 2024. [Human vs. machine-like representation in chatbot mental health counseling: the serial mediation of psychological distance and trust on compliance intention](#). *Current Psychology*, 43(5):4352–4363.
- Soya Park and Chinmay Kulkarni. 2023. Thinking assistants: Llm-based conversational assistants that help users think by asking rather than answering. *arXiv preprint arXiv:2312.06024*.
- Baolin Peng, Chunyuan Li, Pengcheng He, Michel Galley, and Jianfeng Gao. 2023. Instruction tuning with gpt-4. *arXiv preprint arXiv:2304.03277*.
- Paul R Pintrich. 1991. A manual for the use of the motivated strategies for learning questionnaire.
- Sherry Ruan, Liwei Jiang, Justin Xu, Bryce Joe-Kun Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth L Murnane, Emma Brunskill, and James A Landay. 2019. [Quizbot: A dialogue-based adaptive learning system for factual knowledge](#). In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–13.
- Joni Salminen, Soon-Gyo Jung, Johanne Medina, Kholoud Aldous, Jinan Azem, Waleed Akhtar, Essi Häyhänen, and Bernard J Jansen. 2024. [Communication design for an educational ai chatbot: Analyzing cipherbot's communication style and challenges](#). In *Proceedings of the 27th International Academic Mindtrek Conference*, pages 176–187.
- Daniel L Schacter and Karl K Szpunar. 2015. [Enhancing attention and memory during video-recorded lectures](#). *Scholarship of Teaching and Learning in Psychology*, 1(1):60.
- Emanuel A Schegloff. 1968. [Sequencing in conversational openings 1](#). *American anthropologist*, 70(6):1075–1095.
- Anuschka Schmitt, Thiemo Wambsganss, and Jan Marco Leimeister. 2022. [Conversational agents for information retrieval in the education domain: A user-centered design investigation](#). *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–22.
- Noah L Schroeder and Olusola O Adesope. 2014. A systematic review of pedagogical agents' persona, motivation, and cognitive load implications for learners. *Journal of Research on Technology in Education*, 46(3):229–251.
- John R Searle. 1969. *Speech acts: An essay in the philosophy of language*, volume 626. Cambridge university press.
- Hyungyu Shin, Eun-Young Ko, Joseph Jay Williams, and Juho Kim. 2018. [Understanding the effect of in-video prompting on learners and instructors](#). In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–12, New York, NY, USA. Association for Computing Machinery.
- Xiangru Tang, Anni Zou, Zhuosheng Zhang, Ziming Li, Yilun Zhao, Xingyao Zhang, Arman Cohan, and Mark Gerstein. 2024. [MedAgents: Large language models as collaborators for zero-shot medical reasoning](#). In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 599–621, Bangkok, Thailand. Association for Computational Linguistics.
- Deborah Tannen. 2005. *Conversational style: Analyzing talk among friends*. Oxford University Press.
- Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Wei-Lin Chen, Chao-Wei Huang, Yu Meng, and Yun-Nung Chen. 2024. [Two tales of persona in LLMs: A survey of role-playing and personalization](#). In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 16612–16631, Miami, Florida, USA. Association for Computational Linguistics.
- Lev S Vygotsky. 1978. Mind in society: The development of higher mental processes (e. rice, ed. & trans.).
- Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. [Towards mutual theory of mind in human-ai interaction: How language reflects what students perceive about a virtual teaching assistant](#). In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–14.
- Tianjia Wang, Tong Wu, Huayi Liu, Chris Brown, and Yan Chen. 2025. Generative co-learners: Enhancing cognitive and social presence of students in asynchronous learning with generative ai. *Proceedings of the ACM on Human-Computer Interaction*, 9(1):1–24.
- Max Welz and Andreas Alfons. 2023. When respondents don't care anymore: Identifying the onset of careless responding. *arXiv preprint arXiv:2303.07167*.
- Joseph J Williams and Tania Lombrozo. 2010. [The role of explanation in discovery and generalization: Evidence from category learning](#). *Cognitive science*, 34(5):776–806.

- Joseph Jay Williams, Tania Lombrozo, Anne Hsu, Bernd Huber, and Juho Kim. 2016. [Revising learner misconceptions without feedback: Prompting for reflection on anomalies](#). In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 470–474, New York, NY, USA. Association for Computing Machinery.
- Rainer Winkler, Sebastian Hobert, Antti Salovaara, Matthias Söllner, and Jan Marco Leimeister. 2020. [Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent](#). In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–14.
- David Wood, Jerome S Bruner, and Gail Ross. 1976. [The role of tutoring in problem solving](#). *Journal of child psychology and psychiatry*, 17(2):89–100.
- Junchao Wu, Shu Yang, Runzhe Zhan, Yulin Yuan, Derek F. Wong, and Lidia S. Chao. 2023. [A survey on llm-generated text detection: Necessity, methods, and future directions](#). *Preprint*, arXiv:2310.14724.
- Wenquan Wu, Zhen Guo, Xiangyang Zhou, Hua Wu, Xiyuan Zhang, Rongzhong Lian, and Haifeng Wang. 2019. [Proactive human-machine conversation with explicit conversation goal](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3794–3804, Florence, Italy. Association for Computational Linguistics.
- Shaoke Zhang, Hao Jiang, and John M Carroll. 2011. [Integrating online and offline community through facebook](#). In *2011 International Conference on Collaboration Technologies and Systems (CTS)*, pages 569–578. IEEE.
- Peixiang Zhong, Chen Zhang, Hao Wang, Yong Liu, and Chunyan Miao. 2020. [Towards persona-based empathetic conversational models](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6556–6566, Online. Association for Computational Linguistics.

A Additional Learning Sciences Background About In-Video Prompts and Chatbots in Education Contexts

Online learners who passively watch videos alone may struggle to engage deeply with the content, construct knowledge, and avoid distractions (Fang et al., 2022). Research in learning sciences highlights the pedagogical benefits of encouraging learners to articulate their understanding of concepts and facts in their own words (Chi et al., 1994; Aleven et al., 2004). Self-explanation is shown to be effective in facilitating learners to integrate new information into existing knowledge, and learners who are explicitly asked to self-explain are more likely to achieve a correct mental model of a learned concept (Chi et al., 1994). Furthermore, the process of self-explanation helps learners contextualize individual facts within a broader underlying pattern, thereby promoting the generalization of knowledge (Williams and Lombrozo, 2010). Randomized experiments showed that prompting learners with prior misconceptions to explain anomalies (i.e., facts that contradict their existing beliefs) could drive them to revise incorrect beliefs, and this strategy is more effective when learners explain multiple anomalies rather than a single anomaly (Williams et al., 2016).

B The EduPrompt Interface

Figure 3 shows the interface of EduPrompt, where a learning video and a chatbox are placed side-by-side. The video is automatically paused at the end of each segment, and “Alex” sends out a message to initiate a conversation based on the content of the segment that the learner just finished watching.

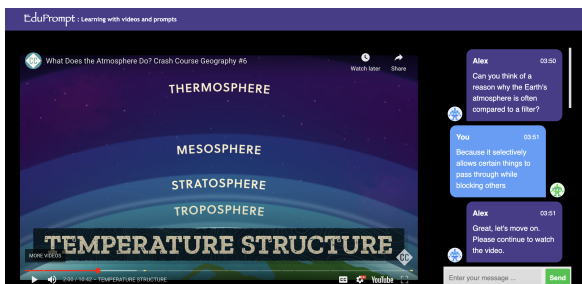


Figure 3: EduPrompt interface. A learning video and a chatbox is presented side by side.

C System Prompt Template

The following template was used as the system prompt to GPT-3.5 to generate Alex’s conversation initiation messages. The template has been iteratively tested and improved to optimize the quality of message generation:

Imagine you are a { teacher || learner } of a recorded lecture video. A { learner || peer learner } is now watching the video. They have just finished watching a segment, and the transcript of the segment is: "{ transcript segment }". You are asked to send { this learner || your peer } a message (to be sent through a chat box), aiming at prompting them to think and learn. Note that this will be a one-off communication and you are not expected to respond to the learner’s reply.

The message must be: (1) A { question || statement } based on the segment’s content. (2) Using a { teacher’s tone || learner’s tone }. Through the message, it should be easy to tell that you are { a teacher, instead of a peer learner || a peer learner, instead of a teacher }. (3) Less than 50 words.

D Additional Details About the Manipulation Check and Examples of LLM-Generated Messages

In the qualitative manipulation check, we verified that GPT-3.5 could closely follow instructions and generate messages that aligned well with the persona and the message type specified in our system prompts. For instance, when asked to send questions, “Alex” as a teacher typically posed straightforward, direct questions, whereas “Alex” as a peer often framed their questions in a more casual, conversational tone. Similarly, a “teacher’s” statement usually combines a summary of factual knowledge with direct instructions for reflection, while a “peer’s” statement typically expresses thoughts and feelings about the video segment and encourages learners to think deeply. Occasionally, Alex’s statement messages, when adopting the role of a peer, begin with a rhetorical question (e.g., “Hey, isn’t it fascinating how ... ?”). Despite this opening, the

message's primary function remains a declarative statement.

Table 1 shows examples of conversation initiation messages in each condition.

E Details About the Study Recruitment

E.1 Main Study

We recruited participants from Prolific⁷, an online recruitment platform commonly used for human subjects research. To be eligible for the study, participants should be (1) at least 18 years old; (2) currently living in the United States; (3) proficient in English; (4) have some experience with watching educational videos online (e.g., watching tutorial videos or popular science videos on YouTube or watching recorded class videos on Coursera, etc.). In addition, only Prolific users who have a higher than 95% of task approval rate can participate. While participants were not required to respond to all of Alex's messages, they were asked to send at least 3 readable messages (i.e., not a random series of characters) as proof of work to pass the attention check. In total, 48 participants completed the study, and we removed data collected from 5 of them who failed attention checks. The remaining 43 participants were paid \$6, which is more than Prolific's minimum payment (\$4.67 per 35 minutes) and adequate based on participants' country of residence. Data was anonymized before analysis was performed.

Among the 43 participants, 16 were male and 27 were female. 5 of them are high school graduates (or equivalent), 15 have some college or associate degree, 13 have a Bachelor's degree (or equivalent), and 10 have a Master's degree (or equivalent). These participants are highly familiar with learning through online videos (e.g., through Youtube, Coursera, etc.); on a scale of 1-7 (with 7 being very familiar), their average self-reported rating is 6.14 ($med = 6$). They are moderately knowledgeable about the learning content before participating in the study; on a scale of 1-7 (with 7 being very knowledgeable), their average rating is 3.93 ($med = 4$). Note that while many individuals may have encountered the concepts mentioned in the introductory-level videos in their daily lives, they typically lack in-depth knowledge of the details that the videos dive into.

E.2 Follow-Up Study

The eligibility criteria for the follow-up study are the same as those for the main study. There were 10 males and 12 females who participated in the follow-up study. Self-reported education background showed that 4 of them are high school graduates (or equivalent), 1 has some college or associate degree, 13 have a Bachelor's degree (or equivalent), and 3 have a Master's degree (or equivalent), and 1 prefers not to say. Similar to main study participants, they are very familiar with online video-based learning, with an average rating of 6.59 out of 7 ($med = 7$). Their average rating of knowledge about the video content is 4.14 out of 7 ($med = 4.5$).

F Survey Instrument

Table 2 shows all survey questions that we used to collect learners' quantitative feedback. Emotional engagement questions were adapted from Burch et al. (2015), cognitive engagement questions were adapted from Burch et al. (2015), self-efficacy of learning questions were adapted from Pintrich (1991), perceived distraction and cognitive load questions were adapted from Hart and Staveland (1988), and questions about perceived rapport with the chatbot were adapted from Zhang et al. (2011). These questions were measured using 7-point Likert-scale items. Participants also rated the chatbot's anthropomorphism, intelligence, and likeability using 5-point Likert-scale questions adapted from Wang et al. (2021). We slightly modified the established survey items to make sure that the survey questions are (1) suitable for our study context and (2) keep the survey concise to avoid participants losing focus during a lengthy study, considering that long surveys often lead to careless responding and threaten internal validity (Welz and Alfons, 2023). For example, when measuring anthropomorphism, we removed Wang et al. (2021)'s question regarding whether the chatbot responded rigidly or elegantly, as this aspect is not relevant to our study context or research questions.

We developed the following open-ended survey questions to collect learners' qualitative feedback. These questions have been tested in several pilot studies to ensure their clarity:

- Why do you frequently/rarely message Alex? Please explain.
- Comparing your learning experience in this

⁷<https://www.prolific.com/> Accessed in 2026.

	Teacher	Peer
Question	<i>How does deep processing differ from shallow processing in terms of encoding information? What strategies can you apply to enhance retention and meaningful connection in your learning process?</i>	<i>Hey there! Have you ever thought about how our memories are processed and stored in our brains? How do you think the concept of short-term and long-term memory impacts our daily lives?</i>
Statement	<i>Understanding implicit processes in memory storage is key to how we learn and recall information effortlessly. Reflect on how your brain automatically processes information before consciously realizing it.</i>	<i>Hey, the way our memories shape who we are is deep. It's like they hold a piece of us that without, we might feel lost. Let's think more on how memory affects our identity.</i>

Table 1: Examples of messages with different personas and message types sent by Alex, the chatbot that was integrated into EduPrompt.

task and your previous online video learning experience, what differences did you perceive? Which one do you like better, and why? Please be very specific.

- How would you expect the experience to differ when learning with a much longer video (e.g., instead of watching a 10-minute video, you'll be watching a 1- to 2-hour-long video) using this tool? How might you adjust your viewing behavior or utilize features of the interface to optimize your experience?
- Suppose you are asked to learn through a series of classes (instead of watching a single video as you just did) by watching online videos. Will you be more/less willing to learn online if you are provided with this tool, compared to the normal way of online video watching (i.e., watching without Alex)? Why?
- How was your experience of chatting with Alex? How did your Alex help or hinder your learning? Please explain in detail.
- (Optional) Thank you again for participating in our study. Is there anything else you want us to know (regarding the study, the tool, etc.)? We sincerely appreciate your feedback!

G Deploying a Learning Companion Chatbot to Support Diverse Learning Scenarios.

While our user study only included one short educational video as the learning material, when we deploy conversation-based learning companion systems like EduPrompt at scale, we need to take diverse types of video-based learning into considera-

tion. For example, when learners are learning with longer videos, the chatbot should limit the number of video segments/messages (e.g., by merging several short segments into one) to avoid learners being interrupted too frequently; frequent interruptions and conversations may lead to distraction, excessive cognitive load, and significantly extend the time needed for finishing video-watching. Overall, we anticipate that thoughtfully designed chatbots that proactively initiate conversations with learners will demonstrate greater benefits when integrated into longer learning sessions or those involving multiple videos. In these scenarios, chatbots can help address the challenges of passive learning, such as distraction, mind-wandering, and information retention issues, which are more significant in extended learning activities. This expectation is supported by learners' comments when they were asked to envision using EduPrompt in prolonged learning sessions.

<i>Measurements</i>	<i>Survey Items</i>
Emotional engagement (Strongly disagree - Strongly agree)	I am enthusiastic about learning using this tool/website. I feel energetic when I am learning using this tool/website. I am excited about learning using this tool/website.
Cognitive engagement (Strongly disagree - Strongly agree)	When I am learning using this tool/website, my mind is focused on video content. When I am learning using this tool/website, I am distracted from video content. When I am learning using this tool/website, I am absorbed by video content.
Self-efficacy (Strongly disagree - Strongly agree)	I believe I will receive an excellent grade if I'm tested based on the content of the video. I'm confident I can learn the basic concepts taught in the video. I believe I will receive a very low grade in this class.
Perceived distraction (Strongly disagree - Strongly agree)	I can concentrate on learning when chatting with Alex, my learning companion. I feel distracted from the learning when chatting with Alex, my learning companion.
Cognitive load (Very low - Very high)	How mentally demanding was the task? How hurried or rushed was the pace of the task? How insecure, discouraged, irritated, stressed, and annoyed were you? I feel attached to Alex.
Perceived rapport (Strongly disagree - Strongly agree)	I maintain a very close social relationship with Alex. I spent a lot of time interacting with Alex. I trust Alex a lot. I believe Alex is helpful.
Chatbot's anthropomorphism	Alex was fake – natural. Alex was artificial – lifelike. Alex was machinelike – humanlike. Alex was unconscious – conscious.
Chatbot's intelligence	Alex was unintelligent – intelligent. Alex was foolish – sensible. Alex was ignorant – knowledgeable. Alex was irresponsible – responsible. Alex was incompetent – competent.
Chatbot's likeability	Alex was unkind – kind. Alex was dislikeable – likeable. Alex was awful – nice. Alex was unfriendly – friendly. Alex was unpleasant – pleasant.

Table 2: Survey items for collecting quantitative data from EduPrompt users (Likert-scale).