

1 **Generative Co-Learners: Enhancing Cognitive and Social Presence of Students**
2 **in Asynchronous Learning with Generative AI**
3

4 TIANJIA WANG, Virginia Tech, USA
5

6 TONG WU, Virginia Tech, USA
7

8 HUAYI LIU, Virginia Tech, USA
9

10 CHRIS BROWN, Virginia Tech, USA
11

12 YAN CHEN, Virginia Tech, USA
13

14 Cognitive presence and social presence are crucial for a comprehensive learning experience. Despite the flexibility of asynchronous
15 learning environments to accommodate individual schedules, the inherent constraints of asynchronous environments make augmenting
16 cognitive and social presence particularly challenging. Students often face challenges such as a lack of timely feedback and support, a
17 lack of non-verbal cues in communication, and a sense of isolation. To address this challenge, this paper introduces Generative Co-
18 Learners, a system designed to leverage generative AI powered agents, simulating co-learners supporting multimodal interactions, to
19 improve cognitive and social presence in asynchronous learning environments. We conducted a study involving 12 student participants
20 who used our system to engage with online programming tutorials to assess the system's effectiveness. The results show that by
21 implementing features to support textual and visual communication and simulate an interactive learning environment with generative
22 agents, our system enhances the cognitive and social presence in the asynchronous learning environment. These results suggest
23 the potential to use generative AI to support students learning at scale and transform asynchronous learning into a more inclusive,
24 engaging, and efficacious educational approach.
25

26 CCS Concepts: • **Human-centered computing** → **Interactive systems and tools**; • **Applied computing** → **E-learning; Interactive**

27 learning environments; Distance learning; Collaborative learning.

28 Additional Key Words and Phrases: Asynchronous learning, Generative AI, Cognitive Presence, Social Presence
29

30 **ACM Reference Format:**

31 Tianjia Wang, Tong Wu, Huayi Liu, Chris Brown, and Yan Chen. 2018. Generative Co-Learners: Enhancing Cognitive and Social
32 Presence of Students in Asynchronous Learning with Generative AI. In . ACM, New York, NY, USA, 23 pages. <https://doi.org/XXXXXX>.
33 XXXXXX
34

35 **1 INTRODUCTION**
36

37 Asynchronous learning, which involves the consumption of pre-recorded educational materials and activities that do
38 not require simultaneous engagement from participants, is an increasingly popular mode of study [22]. In asynchronous
39 courses, materials can include various formats such as slides, pre-recorded lectures, and interactions between study
40 participants may occur via computer-supported methods, such as email, blogs, and discussion boards. The enhancement
41 of cognitive and social presence in asynchronous environments poses a significant challenge due to the intrinsic
42 constraints associated with digital communication mediums, alongside issues such as lack of real-time feedback, limited
43

44

45 Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not
46 made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components
47 of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on
48 servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

49 © 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

50 Manuscript submitted to ACM
51

interaction with the instructor and other students, and low self-efficacy [15]. Learners engaged in asynchronous courses or self-directed study through tutorial videos often experience a diminished sense of cognitive and social presence compared to their counterparts in traditional, in-person educational settings [6, 19, 27], and the lack of cognitive and social presence can lead to suboptimal learning outcomes [2, 20, 21, 30]. Also, students from underrepresented backgrounds often do not have the same social capital or peer cohorts early in their educational journey to help them counteract the effects of isolation [51].

Given these challenges, the need to enhance both cognitive and social presence in asynchronous learning environments becomes crucial. Cognitive presence is the ability of students to construct and confirm meaning through reflection [18]. Social presence involves the level of awareness of another person in an interaction and the resulting level of appreciation of an interpersonal relationship [45, 54]. Both cognitive and social presence have been demonstrated to be positively correlated with the quality of the learning experience [19, 47]. Enhancing cognitive and social presence has been shown to effectively boost work motivation, encourage problem-solving skills, and foster a sense of competition [7, 9, 31].

The advent of advanced generative AI presents novel opportunities to improve cognitive and social presence in asynchronous learning environments. Prior studies have demonstrated the potential of generative agents to mimic human behavior within sandbox gaming [37]. Additionally, different studies have explored the use of generative models in interactive training programs designed for teaching assistants [34]. However, previous studies have been limited to interaction between generative agents in sandbox environments and focused on user interactions via text-based conversations. In this work, we introduce Generative Co-Learners (GCL), a system that leverages multimodal generative agents to act as co-learners in asynchronous study environments to improve cognitive and social presence. In GCL, the co-learners simulate studying alongside users by sharing both screen content and a webcam view that displays visible actions. The system leverages generative AI supporting efficient and interactive exploration of learning materials, facilitating critical and collaborative discourse, and fostering deeper learning engagement by simulating realistic social interactions. Our design incorporates cognitive presence and the three social presence factors, including social awareness, social interaction, and group cohesion [35, 40, 47, 54].

The effectiveness of GCL was assessed through a user study with 12 student participants. Our results indicate that GCL effectively enhances cognitive presence by promoting high user engagement with system features that support the exploration of course materials and facilitate critical discourse with co-learners. Additionally, the system enhances social presence, as shown by higher participant ratings compared to a baseline system, with qualitative feedback underscoring more engaging interactions in collaborative learning. Compared to previous work, which showed that generative agents can effectively simulate human behavior in interactive sandbox environments, we extend this approach to real-world educational settings, where the generative co-learners continuously perceive input from the learning environment and generate believable behavior to interact with real-world users. Our contributions are threefold:

- We propose Generative Co-Learners, an innovative system powered by generative AI powered agents to enhance cognitive and social presence in asynchronous learning environments.
- Through a preliminary user study, we quantitatively and qualitatively demonstrate that GCL enhances cognitive and social presence.
- Our findings offer insights into the design and implementation of using generative AI to support students' learning in asynchronous environments, providing guidance for future research in this domain.

105 **2 BACKGROUND AND RELATED WORK**

106

107 **2.1 Cognitive and Social Presence in Asynchronous Learning**

108 Asynchronous learning has gained significant popularity in educational institutions, particularly during and after the
 109 pandemic period [15]. In asynchronous courses, classes are not conducted in real-time. Instead, content and assignments
 110 are provided in various formats through asynchronous learning tools, allowing students to access materials and complete
 111 their coursework and exams within a specified timeframe. The most common method is for instructors to provide pre-
 112 recorded videos or slides. Fabriz's [15] study on the impact of online teaching and learning settings in higher education
 113 on students' learning experience during COVID-19 explored how these synchronous and asynchronous settings affected
 114 student experiences and outcomes. The findings suggest students in synchronous settings reported more peer-centered
 115 activities, such as feedback, indicating a higher level of social interaction and support for their psychological needs.
 116 In contrast, students in mostly asynchronous settings experienced fewer of these interactions. Garrison's [19] article
 117 underscores the pivotal role of a structured community of inquiry (CoI) framework that encompasses cognitive, social,
 118 and teaching presences as fundamental components for facilitating meaningful online educational experiences. The
 119 work suggests that through strategic facilitation and the leveraging of technology, online education can achieve levels
 120 of critical inquiry and student engagement traditionally associated with face-to-face learning environments.
 121

122

123 *Cognitive Presence.* Cognitive presence is defined as the extent to which learners are able to construct and confirm
 124 meaning through sustained reflection and discourse in CoI [19]. It encompasses four phases of critical inquiry: triggering
 125 event, exploration, integration, and resolution. In asynchronous learning, the triggering event may occur as learners
 126 explore study materials and encounter issues that necessitate further inquiry [3]. During the exploration phase, learners
 127 can examine problems individually and collectively to facilitate reflection and discourse. This leads to the construction
 128 of meaning in the integration phase, and ultimately to the synthesis of findings, transforming abstract knowledge into
 129 practical solutions in the resolution phase. However, existing methods often prove inadequate for effectively supporting
 130 the exploration and integration phases in asynchronous learning environments. These phases require learners to deeply
 131 engage with the material and collaboratively construct knowledge, which can be challenging without effective methods
 132 for seeking help, discussing with peers, and receiving immediate feedback. Prior work has explored ways to make
 133 remote help-seeking paradigms more efficient through novel communication techniques [8, 10, 11]. However, they still
 134 required human helpers in the loop which is not scalable. Our system aims to enhance cognitive presence by providing
 135 clearer, more immediate ways for students to interact with study materials and by leveraging generative co-learners to
 136 offer real-time feedback and facilitate discussions.
 137

138

139 *Social Presence.* Short's paper first introduces the concept of "social presence" as the perceptibility of another
 140 individual in communication and the resulting prominence of interpersonal relationships [45]. In CoI, Garrison defines
 141 social presence as the ability of participants in the Community of Inquiry to project their personal characteristics
 142 into the community, thereby presenting themselves to the other participants as "real people". Subsequent research
 143 has sought to refine this concept within online learning contexts. For instance, Weidlich described social presence
 144 as the perception of others as "real" in communication, emphasizing the subjective sensation of being in a shared,
 145 technologically mediated space [54]. Öztok further narrowed the definition to the sense of "being there, together", in
 146 the absence of physical co-presence [36]. Where in CoI, group cohesion is important for fostering a sense of belonging
 147 and facilitating critical inquiry, as well as enhancing the quality of discourse. When students perceive themselves as
 148 members of a cohesive group rather than as isolated individuals, they are more likely to share personal meanings
 149

and engage more deeply in the learning process. This sense of unity is crucial for the effective exchange of ideas and collaborative learning. Additionally, Weidlich endeavored to enhance the SIPS (Sociability, Social Interaction, Social Presence, Social Space) model by differentiating social presence from other social dimensions relevant to online education [53]. Moore's seminal work on the differentiation within online and distance learning posits that social presence is essentially about the interactions among students, characterizing it as a facet of social interaction [35]. According to Weidlich, social interaction not only initiates the formation of impressions of others but also serves as a vital precursor to the development of social presence [53]. Research by Swan and Shih has shown a significant correlation between group cohesion, social presence, and perceived learning outcomes [47]. Similarly, Ratan outlined peer social presence in online learning environments as the students' awareness of each other, comfort in communication, and a sense of togetherness [40].

Efforts to enhance social presence in digital learning environments have employed technologies like discussion forums, online live chats, and screen sharing tools, aiming to incorporate visual cues that facilitate engagement [7, 29]. For example, Fang's article examines how structured note-taking systems enhance comprehension and engagement for video-based learners in both individual and social learning contexts [16]. However, these technologies have not been able to replicate the depth of social presence inherent in traditional face-to-face learning environments. Traditional learning settings offer rich, dynamic interactions, where learners benefit directly from engaging with peers and observing their behaviors, fostering an immersive social presence [40]. This gap highlights the need for further research to develop solutions that can more effectively bridge the divide between digital and in-person learning experiences. Drawing from previous research, this paper aims to enhance social presence in asynchronous learning environments by leveraging multimodal generative agents to act as co-learners to improve social awareness, social interaction and group cohesion.

2.2 Generative AI in Education

The existing literature demonstrates the utility of AI tools powered by Large Language Models (LLMs) in the realm of higher education [24, 46]. Wang et al. explored the capabilities of ChatGPT, a popular LLM-based chatbot, in computer science education, demonstrating its ability to successfully solve computing problems across levels and topics—outlining problem modification methods and instructor concerts [52]. Sarsa's article investigates the use of OpenAI Codex, a large language model, for generating programming exercises and code explanations, finding that most content created is novel and sensible and suggests significant potential for these models in aiding programming education [44]. Markel's work presents GPTEach, a tool using GPT-simulated students for interactive teacher training. It shows that it allows novice teaching assistants to practice teaching strategies flexibly and was preferred over traditional rule-based dialogue systems for training [34]. These studies demonstrate the effectiveness of LLM-driven AI tools in augmenting students' learning experiences. Our research also contributes to the evolving landscape of educational technologies by further exploring the application of using the state-of-the-art AI model to support student learning.

In recent advancements, LLMs like GPT-4 [1] have shown remarkable capabilities in comprehending complex language inputs and generating coherent, contextually relevant responses [5, 14]. These models, known for outperforming humans in specific tasks, particularly excel in areas requiring extensive knowledge or processing of vast information. This proficiency is evident in closed-book question-answering tasks, where LLMs surpass human performance on datasets such as NaturalQuestions, WebQuestions, and TriviaQA [57]. Their ability stems from training on billions of tokens, encompassing a wide range of real-world knowledge. A study by Kosinski using 40 false-belief tasks, a benchmark for testing Theory of Mind (ToM) in humans, revealed that while earlier and smaller LLMs lacked ToM abilities, GPT-4

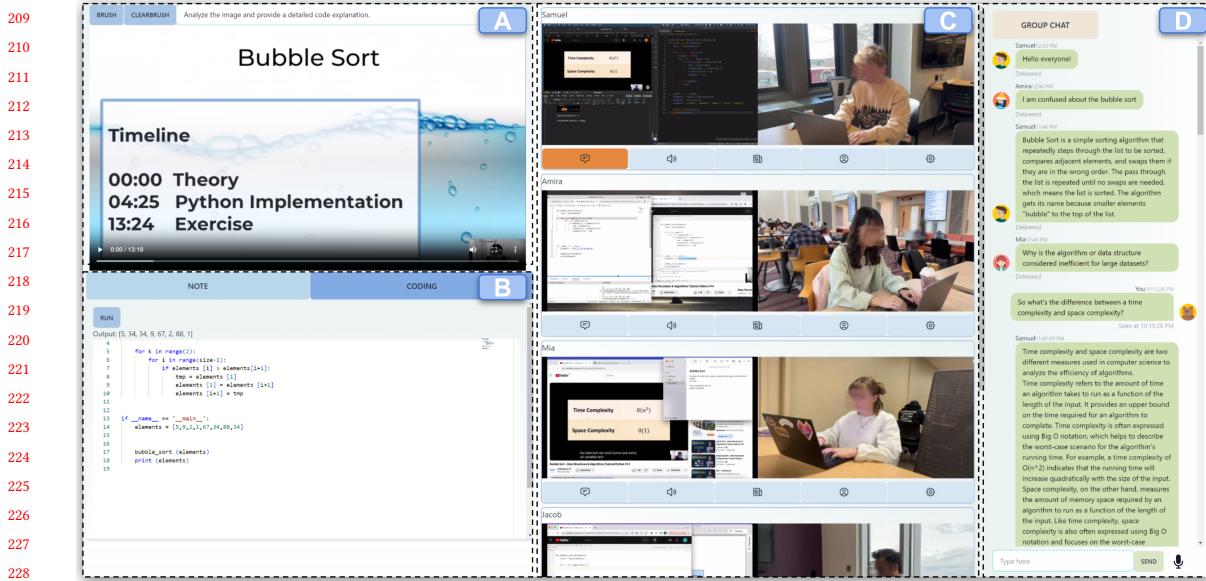


Fig. 1. An overview of GCL user interface. There are four panels: (A) a main video panel for displaying the learning content; (B) a function panel with a text and code editor; (C) a co-learners panel displaying the video and screen share of simulated co-learners in asynchronous learning environments powered by generative AI; and (D) a chat panel for users to communicate with co-learners.

showed comparable performance to six-year-old children. The result suggests that ToM might be an emergent property in advanced AI models [28].

Building on the robust capabilities of LLMs, various studies have explored their application in simulating human behavior and performing software development tasks. Park's research employed LLM-powered generative agents to mimic human interactive behaviors in the Sim's game, using a memory-inclusive generative agent architecture [37]. This work provides insights into how LLMs can be configured and prompted for human-like actions such as planning, perception, reflection, and action. Qian's article introduces CHATDEV, a framework that uses generative agents in a simulated waterfall software development process [39]. In our research, we introduce GCL—a system that employs generative AI-powered agents as co-learners, supporting learners in exploring study materials, monitoring the environment to generate believable actions, and engaging in multimodal interactions with users to promote computer-supported collaborative learning in asynchronous learning environments.

3 GENERATIVE CO-LEARNERS

3.1 Design Goals

Drawing from previous research, this paper aims to enhance cognitive and social presence in asynchronous learning environments through generative AI-powered agents with user-friendly interfaces. In an asynchronous learning environment, students may struggle with cognitive presence due to a lack of efficient information exchange and way to help them connect ideas [19]. Further, asynchronous settings may inhibit social presence for students due to limited interpersonal relationships [40, 54], a lack of communication, connection, and interaction between individuals within

the learning environment [35, 53], and a minimal sense of togetherness and meaningful collaborations [40, 47]. Thus, we utilized the following design goals to integrate cognitive and social presence into our system:

- Supporting the *cognitive presence* for the user during the learning progress (DG1)
- Enhancing the *social presence* with the three factors (DG2)
 - Improving the *social awareness* of co-learners (DG2-1),
 - Facilitating *social interaction* between users and co-learners (DG2-2),
 - Enhancing *group cohesion* in the learning environment (DG2-3).

3.2 Motivating Example

Imagine Alice, a student enrolled in an online computer science course. Due to her work commitments and limited access to a local academic community, Alice chooses asynchronous courses, engaging with pre-recorded lectures and coding tutorials at her own pace. Although this mode of study fits her schedule, she often encounters common challenges in such learning environments, including delayed feedback on her learning progress, difficulties in grasping complex topics without an effective way to explore the study materials, and a lack of peer engagement leading to feelings of isolation and reduced motivation.

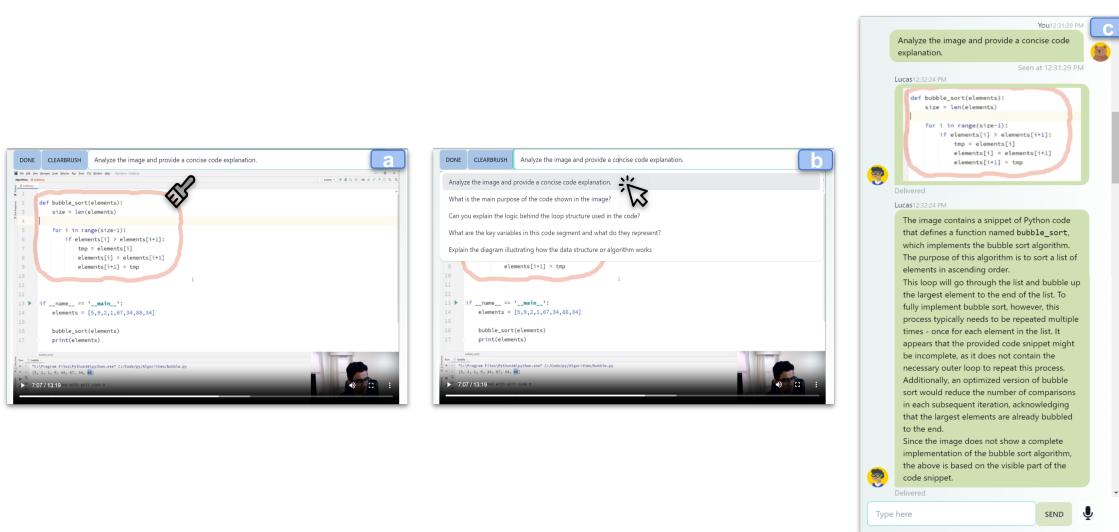
Alice began using the GCL system to study an online asynchronous programming course focused on data structures. Upon entering the system, she noticed other co-learners sharing their screens and webcams, all studying the same course materials. She clicked on each co-learner's profile to learn their names and backgrounds, forming initial impressions about them. As she started watching a pre-recorded lecture video, she became confused about the time complexity of bubble sort. She sought help by sending an audio chat message, "Hi, could you explain why the worst-case time complexity of bubble sort is $O(n^2)$?" to a co-learner. Subsequently, one of the co-learners responded with audio and a message in the chat window, explaining the required number of comparisons and swaps in the worst case with a concrete example, and the screen on the right side was enlarged to show the co-learner pointing at the screen to explain the concepts. She also reviewed the group chat window and found other co-learners actively discussing the space complexity of bubble sort. Feeling that GCL created an active learning environment with a better sense of togetherness, she joined the discussion, gaining a clearer understanding of the time and space complexities of bubble sort.

When she reached the diagram illustrating bubble sort in the video, she became confused about how element swapping works. She clicked the brush button, circled the diagram, and asked, "Can someone explain the concept of swapping elements in this diagram of bubble sort?" A randomly selected co-learner received the message, analyzed the image, and provided a detailed explanation in the group chat about why and how the two adjacent elements are swapped, with the exact numbers shown in the diagram. While learning the theory part of bubble sort, Alice took notes using the note function in GCL but felt she missed some key points. Struggling with her review, she noticed a co-learner actively sharing notes in the chat window that included some points she had missed. Alice then clicked on the notes button and began reading other co-learners' notes to enhance her learning.

After completing the theory section in the video, Alice started working on the actual programming implementation. She followed the tutorial in the lecture video and began writing code using the code editor in GCL. She made an error with the index in a list and was stuck for a while. The GCL detected this, and a co-learner reviewed her code and pointed out the mistake with the list index, helping her progress further. Alice's experience highlights how the GCL system effectively addresses the typical challenges of asynchronous learning by improving cognitive and social presence to foster interactive and supportive collaborative learning.

313 3.3 System Design

314 The user interface of the GCL system is depicted in Figure 6, comprising four primary components: the Main Video
 315 Panel (Figure 6.A), the Function Panel (Figure 6.B), the Co-Learners Panel (Figure 6.C), and the Chat Panel (Figure 6.D).
 316 The Main Video Panel integrates functionalities for activating and clearing the brush tool, alongside a text input box for
 317 posing questions using the brush feature, and displays the principal lecture video. The Function Panel is equipped with
 318 two tabs, enabling users to toggle between note-taking and coding functionalities. The Co-Learners Panel contains six
 319 co-learners, each with a shared screen, an action screen, and five functional buttons for chat, audio chat, notes, profile,
 320 and customization options. The Chat Panel features both group and private chat windows, showcasing chat history,
 321 and includes an input box for users to send messages. The UI for our system incorporates multimodal interactions that
 322 have been shown to enhance learning [58] and engagement [43] in learning environments.
 323



346 Fig. 2. A brush feature enabling users to highlight a specific area in the video and pose related questions to the co-learners.

347 **3.3.1 Main Video Panel.** The main video panel is situated in the top-left corner, featuring a video player dedicated
 348 to delivering educational materials such as lectures and tutorial videos. We recognize the challenges learners may
 349 face when asynchronously engaging with these videos, particularly when encountering difficult concepts but lacking
 350 immediate access to tools for extracting video content or seeking assistance from external resources. For instance,
 351 learners might struggle to comprehend a diagram illustrating a sorting algorithm or a snippet of code presented in the
 352 video, without any efficient ways of obtaining explanations for these elements.

353 To solve the problem, we introduce a *brush feature* aimed at supporting user's cognitive presence (DG1) and interactive
 354 learning (DG2-2) with the video content shown in Figure 2. This feature enables users to select or highlight a specific
 355 area within the video (Figure 2.a) and pose questions related to it, either by choosing from a set of predefined prompts
 356 or entering a custom question in the provided input field (Figure 2.b). This functionality is facilitated by overlaying
 357 a drawing panel on the main video, where the coordinates ($min_x, min_y, max_x, max_y$) of the selected area are tracked.
 358 Upon the user completing the selection and pressing the 'Done' button, the system captures a screenshot of the defined
 359 area and displays it in a separate window for analysis (Figure 2.c). The analysis results, such as a code explanation or
 360 a diagram, are then shared with the co-learner, providing them with the necessary context to understand the highlighted
 361 content. This multimodal interaction allows learners to engage with complex concepts in a more intuitive and
 362 effective manner, ultimately enhancing their overall learning experience.

365 area ($min_x, min_y, max_x, max_y$) and forwards this screenshot along with the user's question to a random co-learner. The
 366 responding co-learner, leveraging the GPT-4 vision¹ model, will analyze the image and question to provide textual
 367 feedback, which is subsequently shared in the group chat window (Figure 2.c).
 368

369 3.3.2 *Function Panel*. The function panel is designed to integrate note-taking and coding tools the learner will need to
 370 engage with educational content. The note-taking tool is designed with a user-friendly interface that enables learners
 371 to write down key points and reflections in real-time as they navigate through the educational material. The coding
 372 tool is tailored for students interacting with programming-related content. It allows users to write and execute codes
 373 within the same environment, offering immediate execution results on their coding exercises.
 374

375 376 3.3.3 *Co-Learners Panel*. In the Co-Learners Panel, our system leverages six generative co-learners, to facilitate
 377 collaborative learning alongside users. These agents are designed to engage in effective communication with the user,
 378 simulating the dynamics of peer-based learning. To achieve this, we initialize each agent with a prompt template that
 379 positions them as a peer learner in an online programming course, along with course information, the concept to be
 380 learned, and the settings of the communication behavior. Each co-learner is equipped with a memory module capable
 381 of storing historical information, enhancing their ability to provide contextually relevant assistance. Upon initialization,
 382 we supply the agents with the full transcript with timestamps of the learning material, enabling them to respond to
 383 content-specific questions and generate notes accordingly.
 384

385 386 For each co-learner, we introduced a shared screen feature, positioned on the left side of each co-learner's panel. By
 387 displaying shared activities, such as watching the same tutorial video, taking notes, and coding, the system simulates
 388 a collaborative learning environment that fosters participation, empathy, and a sense of togetherness (DG2-3). We
 389 recorded a 15-minute session of the shared screen for each co-learner. This session included activities such as watching
 390 the same tutorial video, taking notes, and coding, all aimed at enhancing group cohesion and fostering a positive
 391 learning atmosphere. The shared screen playback continues uninterrupted but pauses automatically to accommodate
 392 break and active actions described in the following paragraphs, and playback resumes once the actions are completed.
 393

394 395 To integrate awareness of co-learners (DG2-1), we introduced an action screen feature, functioning as a virtual
 396 window in the Co-Learner's panel powered by generative AI, to display co-learners' faces and their learning activities.
 397 Inspired by observations of typical behaviors in both offline study environments and "study with me" videos, we
 398 identified 10 passive actions reflective of a learner's study session. Passive actions are actions that occur spontaneously
 399 or without deliberate initiation during a study session. These actions are not triggered by external interactions but arise
 400 from the individual's own volition or as a part of routine behavior. We identified five actions related to study behaviors
 401 (typing, watching, thinking, taking notes, and expressing confusion) and five actions associated with break behaviors
 402 (stretching, rubbing eyes, eating, drinking, and checking the phone). We hired six actors to emulate these actions,
 403 recording their performances to be showcased on the action screen. The typing action is set to play continuously by
 404 default. To simulate a dynamic learning environment, our system uses a scheduler to randomly cycle through study
 405 behaviors and initiate break activities for the co-learners at intervals ranging between 90 to 180 seconds. Upon the
 406 completion of an action, the system automatically reverts to the typing action.
 407

408 409 To facilitate social interaction between users and co-learners (DG2-2), we integrated active actions into the action
 410 screen. We define active actions that are initiated as a response to communication or interaction. These actions are
 411 directly triggered by dialogues, questions, or any form of communication that requires an immediate or specific response.
 412

413 414 415 ¹<https://platform.openai.com/docs/guides/vision>

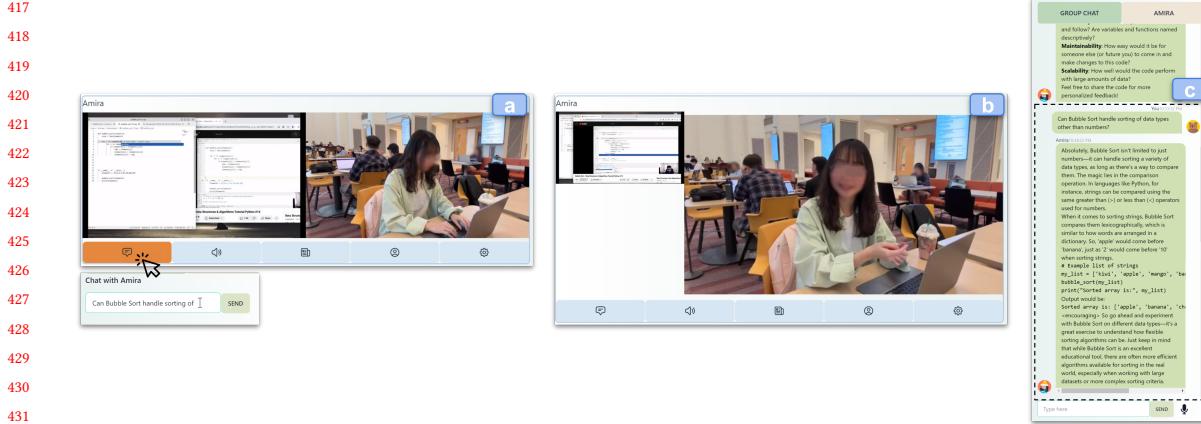


Fig. 3. A text-based chat function that enables users to (a) send text messages to co-learners and ask questions related to the learning material. (b) When the co-learner responds to the user's message, the generative agents select the appropriate active action and display it on an enlarged responsive screen. (c) The text response will be displayed in a private chat window

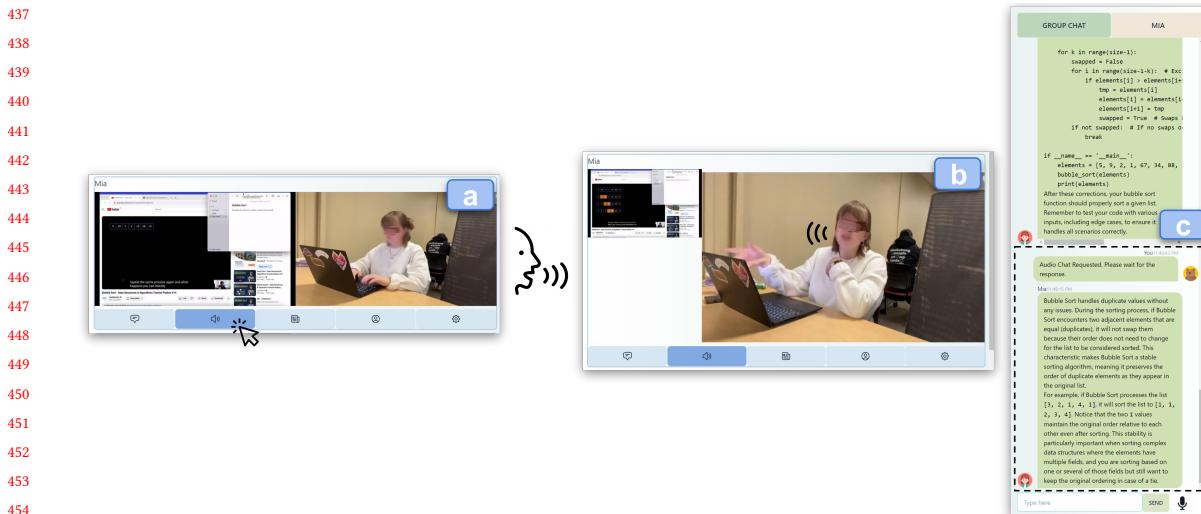


Fig. 4. An audio chat function that enables users to (a) send voice chat messages to co-learners by clicking on the audio chat button. (b) When the co-learner responds to the user's message via audio, the generative agents select the appropriate active action and display it on an enlarged responsive screen. (c) The text from the audio response will be displayed in a private chat window

460
461 Drawing inspiration from behaviors observed by two authors in peer and group study settings, we identified six active
462 actions designed to emulate positive communication behaviors: asking, chatting, encouraging, exciting, explaining, and
463 welcoming. During these actions, co-learners appear to turn towards the user and engage in conversation, with visible
464 mouth movements, body languages, facial expressions and gestures. These actions are performed by our hired actors.
465 Whenever a user initiates active communication with a co-learner, the co-learner will select an appropriate action to
466 display. Each action is divided into three phases: starting, continuing, and ending. The starting phase is triggered when
467

a user begins a conversation, either by sending a text message in the chat window or using the audio chat function, depicting the co-learner turning towards the user. The continuing phase shows the co-learner “speaking” with the action aligned to the context of the interaction, with the duration adjusted to match the length of the co-learner’s text or audio response. For instance, if a user asks a question via audio chat, the co-learner might response with a detailed explanation of the user’s question, and select the “explaining” action. In this scenario, the co-learner would turn to face the user, adopting a facial expression indicative of an explanation, and may gesture toward their screen as if to clarify a point, with the action’s duration tailored to match the length of the audio response.

To support critical discourse (DG1) in cognitive presence and further foster social interaction (DG2-2) we designed our system to allow users to perform multimodal communication with co-learners. This not only includes the brush feature and the action feature we introduced above but also a *text-based chat function* and an *audio chat function* shown in Figure 4. The text chat in Figure 3 enables users to send messages to co-learners and inquire about any questions and concerns during their learning journey.

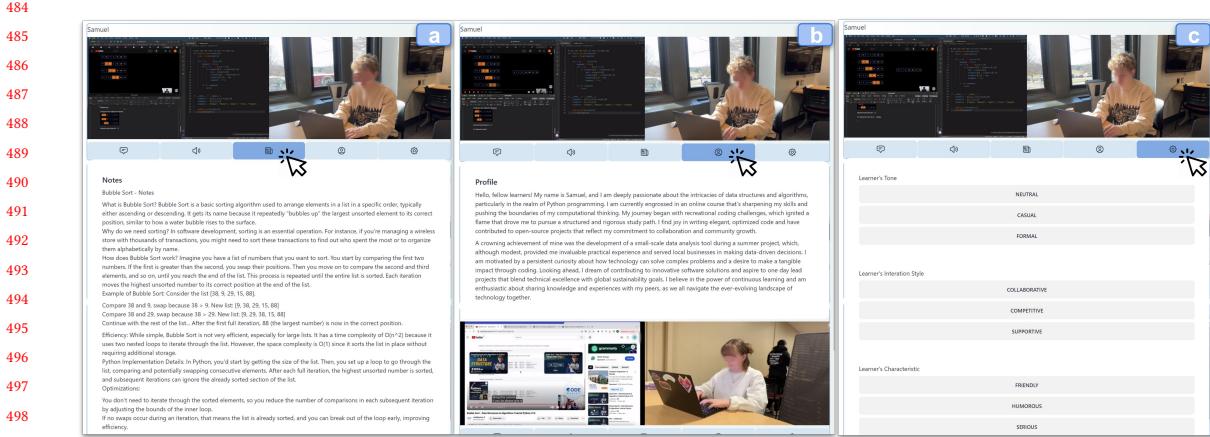


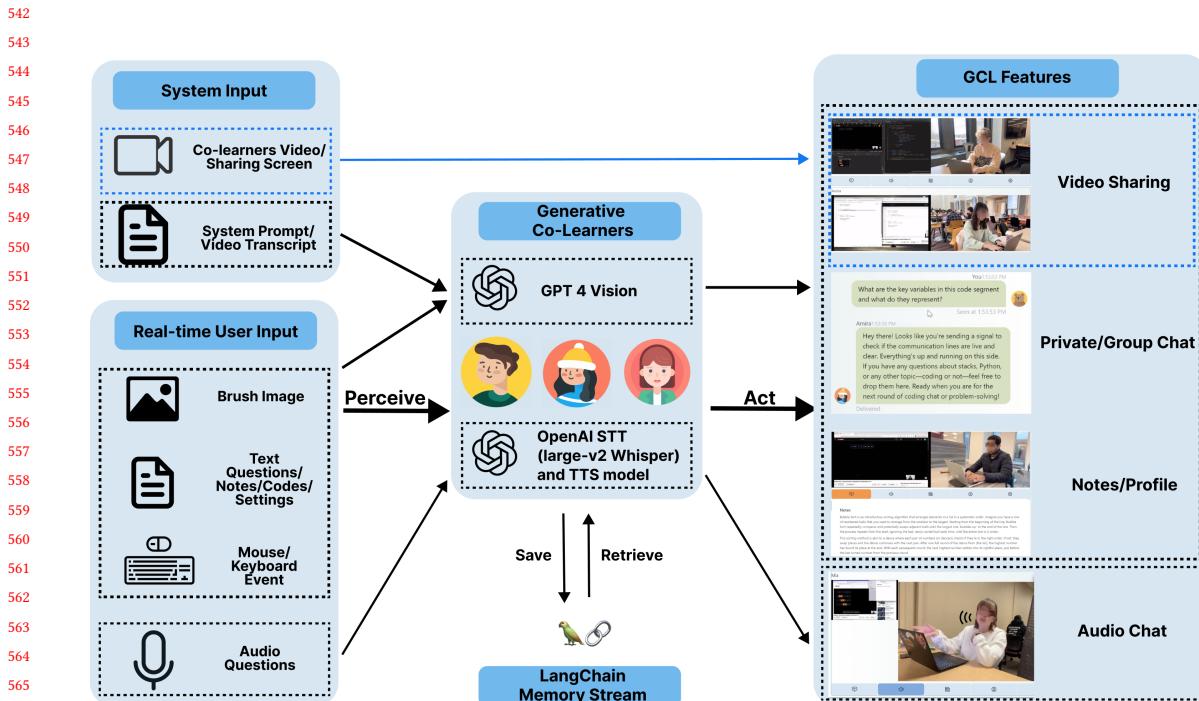
Fig. 5. User can view the co-learner generated (a) notes, (b) profile, and select preferred learner’s tone, interaction style, or characteristic in (c) customization menu

To utilize the audio chat feature, users could press the audio chat button to start recording their message and press it again to end the recording. The voice message is then converted into text using a speech-to-text (STT) model before being sent to the co-learner. In response, the co-learner crafts a text reply, which is subsequently converted into an audio message through a text-to-speech (TTS) model. Both models are provided by OpenAI’s API ², and the TTS supports a range of voice options for the generative agents. Following this, the audio reply is played back, the action video panel enlarges, and an appropriate action chosen by the generative agent is being displayed. Additionally, for reference, the text version of the conversation is displayed in the co-learner’s private chat window. Each co-learner is equipped with notes derived from video content (Figure 5.a), a profile for the self-introduction (Figure 5.b), and a customization feature for configuring tone, interaction style and characteristic (Figure 5.c). The notes could support student’s cognitive presence that aid in comprehending concepts presented in the video tutorials (DG1). Additionally, the profile plays a crucial role in shaping first impressions and facilitating social awareness (DG2-1). The customization feature will allow the user to tailor their learning experience to suit their preferences and needs.

²<https://openai.com/blog/openai-api>

521 Additionally, our system has the features to monitor user activity, including inactivity in mouse movement, note-
 522 taking, and coding, to trigger proactive communication of co-learners and improve students' cognitive presence (DG1,
 523 DG2-2). If a user's mouse is idle for an extended period, the system's scheduler will notify a co-learner who will then
 524 reach out to inquire about the user's progress. Similarly, if the system detects inactivity in taking notes, a co-learner will
 525 be informed and prompted to share notes with the user. For coding activities, if a user does not input new code in the
 526 code editor for more than 60 seconds, a co-learner will be notified to offer code review assistance, aiding in debugging
 527 efforts. All the inactivity time intervals can be adjusted based on a preferred length. These mechanisms are designed to
 528 foster a supportive and interactive learning community.
 529

530
 531 3.3.4 *Chat Panel*. The chat panel primarily features chat windows for both group and private conversations among co-
 532 learners. To foster a sense of togetherness and enable the user and co-learners to participate in meaningful collaboration
 533 (DG2-3), we have implemented a group chat function. This allows co-learners to engage in discussions with one another,
 534 with the added flexibility of joining the chat at any time. This functionality is facilitated by a system scheduler that
 535 forwards a previous message from one co-learner to another within the group chat. Additionally, when a user actively
 536 participates by sending a message in the group chat, the system will randomly select between one to three agents to
 537 provide responses. This approach enables users to encounter a variety of messages, thereby enhancing their learning
 538 experience and social presence through exposure to diverse perspectives.
 539



567 Fig. 6. An overview GCL Workflow. The Generative Co-Learners take initial system input, keep perceiving from real-time user input,
 568 and continuously act on system features. The memory stream saves new information and retrieves historical data.
 569

573 3.4 Implementation

574 Our system is developed as a comprehensive full-stack web application. The architecture includes a front-end constructed using ReactJS³ for the user interface, complemented by a back-end developed with Node.js⁴ to support the
 575 AI-powered generative agents. Communication between the front-end and back-end is facilitated through HTTP requests.
 576 The generative agents are powered by OpenAI's GPT-4 Vision API, and their memory is built using LangChain's
 577 ConversationSummaryBufferMemory⁵. The audio chat feature is implemented by using OpenAI's tts-1 model for
 578 text-to-speech, and whisper-1 model for speech-to-text.⁶ The overview of our system workflow is shown in Figure
 579 6. Initially, the system will use pre-recorded Co-learner videos, shared screen videos and system prompts along with
 580 tutorial video transcripts as input. The videos are controlled by a video scheduler to display the passive actions and
 581 shared screen of co-learners. The system message prompt and video transcript will be the system input when initializing
 582 the generative co-learners. The system message prompt can be found in the Appendix A. During runtime, the generative
 583 co-learners continuously perceive real-time user input from the learning environment, including brush images, chat
 584 messages, notes, codes, chat messages, audio questions, and mouse and keyboard events. Based on the real-time user
 585 input, the generative co-learners will act to interact with the user, such as controlling the video scheduler to display
 586 active actions during interactions, sending chat messages in the private chat or group chat, updating notes and profiles,
 587 and engaging in audio chats with the user. More implementation details of GCL can be found in Appendix B. For testing
 588 and evaluation purposes, both the front-end and back-end were deployed locally on our development machines.
 589

590 4 PRELIMINARY EVALUATION

591 4.1 Study Design

592 *4.1.1 Participant Recruitment.* To recruit actors for the GCL system, we posted a recruitment message on our institute's
 593 social media platforms. The participants were informed that their role would involve acting as co-learners within
 594 the GCL system. They were made aware that videos of their actions would be recorded and displayed in the GCL
 595 system, and that generative AI would be used to power these co-learners' interactions with users. The actors provided
 596 consent for their recordings to be used in the creation of generative AI-powered co-learners. Each recording session
 597 lasted approximately one hour, and participants were compensated \$20 for their time and effort. To evaluate GCL,
 598 we conducted a user study with 12 student participants to test our system in in-person study sessions. We employed
 599 social media platforms and mailing lists within our institute for recruitment. Participants will be required to complete
 600 a screening survey to verify their basic Python skills and prior knowledge of data structures and algorithms. The
 601 participants are informed that they will be working with AI-powered generative co-learners in the GCL system. All
 602 the recruited participants are students with basic Python knowledge who had not previously studied data structures
 603 and algorithms, and the demographic information provided in Table 1. The gender distribution was as follows: 58.3%
 604 (7/12) male, 41.7%(5/12) female, and 8.3% (1/12) non-binary. The average Python experience was approximately 1.42
 605 years, with a median experience of 1 year. Participants' educational levels ranged from freshman undergraduates to
 606 PhD students. Participants were compensated \$20 for their time. The study was approved by the Institutional Review
 607 Board (IRB) at our institution.
 608

609
 610 ³<https://react.dev/>

611 ⁴<https://nodejs.org/>

612 ⁵<https://www.langchain.com/>

613 ⁶<https://platform.openai.com/docs/models/tts>

625 4.1.2 *Learning Content.* We selected two tutorial videos from the codebasics YouTube channel focused on key pro-
 626 gramming concepts as the learning content for our evaluation, one covering stack⁷ and the other describing bubble
 627 sort⁸ in Python. These concepts represent essential topics within data structures and algorithms courses. codebasics⁹ is
 628 an online education platform that offers courses and resources across programming and data analytics. Their YouTube
 629 channel consists of a community of over 1 million subscribers and features over 853 videos related to various topics
 630 in computing, including introductory programming concepts. Each tutorial selected for our evaluation is around 12
 631 minutes long and is presented by the same instructor. The videos were chosen for their high ratings (3.8K and 1.7K
 632 upvotes) and served as the learning material within our system.
 633

634 To measure participant understanding before and after engaging with the material, we designed pre-quizzes to assess
 635 their prior knowledge of the programming concepts and post-quizzes to assess their understanding of each concept.
 636 Each quiz consisted of five multiple-choice questions of similar difficulty, with each correct answer worth one point.
 637 Responses were evaluated by comparing them to a set answer key.
 638

639 4.1.3 *Baseline System.* We implemented two versions of the GCL system for the study: Version A, with interactive
 640 features and AI-generated content disabled to represent a baseline, including the brush feature in the main video panel,
 641 the five functions buttons in the co-learner panel, and the entire chat panel. The co-learners only share the screen
 642 and perform passive action and will not have any interaction with the user. The Version B offers full GCL system
 643 functionality. Each in-person session lasted roughly 60 minutes, during which participants interacted with both system
 644 variants and both learning concepts in a counterbalanced manner to minimize order effects.
 645

646 Table 1. For the user study, we recruited 12 participants with basic Python knowledge who had not previously studied data structures
 647 and algorithms.
 648

PID	Gender Identity	Python Experience	Degree Level
P01	Non-Binary	<1 year	1st year undergraduate
P02	Female	1 year	1st year undergraduate
P03	Male	<1 year	1st year undergraduate
P04	Male	1 year	2nd year undergraduate
P05	Male	3 years	Master
P06	Female	<1 year	Master
P07	Male	1 year	1st year undergraduate
P08	Male	<1 year	4th year undergraduate
P09	Male	<1 year	1st year undergraduate
P10	Female	<1 year	1st year undergraduate
P11	Male	7 years	PhD
P12	Female	1 year	1st year undergraduate

666 4.1.4 *Study Procedure.* The study procedure began with an introductory session during which participants provided
 667 their consent. Subsequently, participants were asked to evaluate the two variants of our system. For each system variant,
 668 participants completed the pre-quiz. Following this, we introduced the system, allowing participants to freely explore
 669 the user interface and system features, and allocated an 18-minute session for participants to use the system to learn
 670 the concepts covered in the tutorial videos.
 671

7⁷<https://www.youtube.com/watch?v=zwb3GmNAtFk>

8⁸<https://www.youtube.com/watch?v=ppmlOUIz4uI>

9⁹<https://codebasics.io/>

677 the concept in a simulated asynchronous learning setting. For the GCL system, we recorded the frequency with which
 678 participants utilized each feature. Upon completing the learning session, participants completed the post-quiz and a
 679 survey to capture their perceptions of the system. The survey was comprised of Likert scale questions regarding whether
 680 they feel the system supports the three factors related to social presence—awareness, social interaction and group
 681 cohesion—and questions for evaluating the overall social presence facilitated by the system and their general satisfaction
 682 with the environment the system provides. After participants completed both learning sessions, we conducted a
 683 10-minute semi-structured follow-up interview to gather insights on the system's notable features, its advantages
 684 and disadvantages, and how the learning experience with GCL compared to previous learning experiences. Our study
 685 instruments, including background and screening survey, pre-quizzes, post-quizzes, post-survey, and post-interview
 686 questions are available online.¹⁰
 687
 688

689 4.2 Results

690 4.2.1 *Can GCL improve cognitive presence?* To investigate the effects of GCL on cognitive presence, we observed
 691 participants' usage of various features of our system. During the study session, the triggering event phase of cognitive
 692 presence occurred when participants expressed their confusion about the content in the tutorial video. When the
 693 triggering event was triggered, we found that participants used shared notes an average of 5.75 times, chat features 4.67
 694 times, and the brush feature 1.17 times to further support their inquiry. The participants perceived the notes provided by
 695 the generative agents as helpful in learning the content of the video, with P01 stating, “*Notes are in-depth and polished*
 696 and go through step-by-step”. The participants found the brush feature very helpful for learning the content of the video,
 697 with P03 saying, “*To point to a spot where you can just visually direct the other co-learners to where your questions are.*
 698 *That saves a lot of time. That's a lot more efficient*”, P10 mentioned, “*Different co-learners can provide different learning*
 699 *suggestions*”, and P04 added, “*I think it's important to take that input from some of the other co-learners, as they may have*
 700 *a more efficient way to write something. There may be a different syntax that they're using*”.

701 The majority of the participants (8/12, 66.66%) believe the AI-generated feedback and response in the communication is
 702 of good quality. Three participants (P08, P11, P12) think the co-learners replies are too long, such as P08 stating the
 703 co-learners response is “*Too wordy, need more concise*”.

704 **Finding:** Our results suggest the system features, including shared notes, multimodal chat, and the brush tool, can
 705 support cognitive presence during participants' learning progress in asynchronous learning environments.

706 4.2.2 *Can GCL enhance social presence?* We explored the capabilities of GCL to improve the three social presence
 707 factors—social awareness, social interaction, and group cohesion—through our survey. According to our findings,
 708 participants ranked the system higher across all the three factors of social presence than the baseline system. The baseline
 709 system achieved average scores of 3, 2.7, 2.8 and 3.7 on a 5-point Likert scale for social awareness, social interaction,
 710 group cohesion, and overall social presence. In contrast, the GCL system received higher average scores of 3.92, 4, 3.92
 711 and 4.3 for the three factors and overall social presence. When we follow a think-aloud process to ask why participants
 712 gave the scores, the majority participants(9/12, 75%) believe the GCL system better supports social awareness, stating
 713 that “*It felt more like a collaborative effort. I can see and talk to interact with these people, and we can learn together, and*
 714 *this provides a better social awareness. For the baseline system, it's a lot of distraction because you're not getting any input*

715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728

¹⁰<https://anonymous.4open.science/r/GenerativeCoLearnersStudy-AE72/>

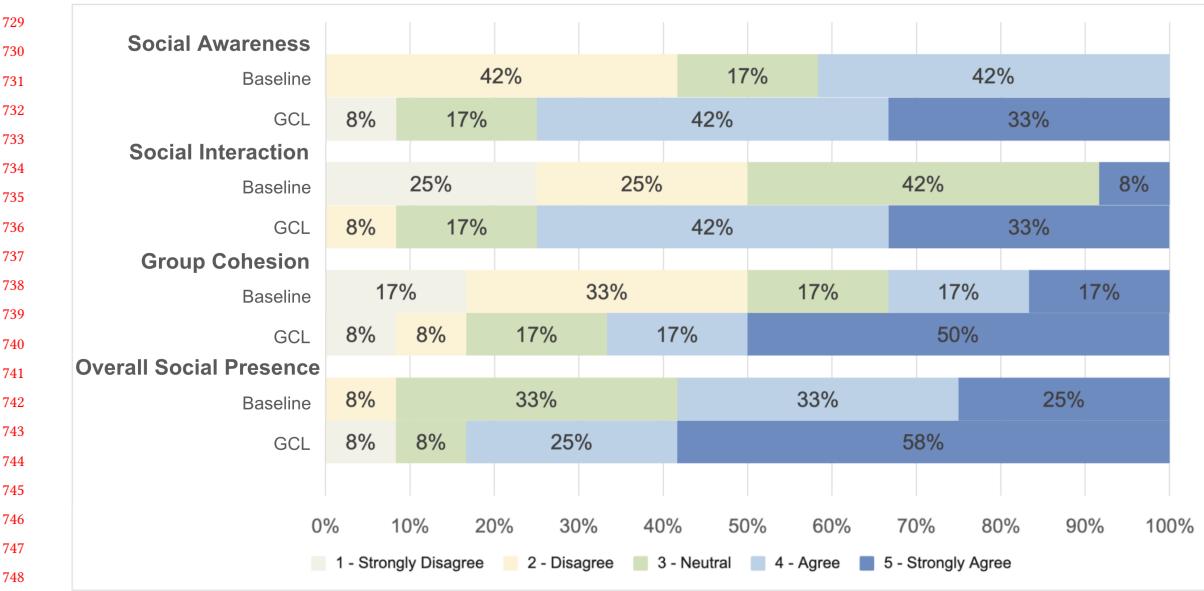


Fig. 7. Participant Ratings on the GCL and baseline system's Enhancement of Social Awareness, Social Interaction, Group Cohesion, and Overall Social Presence, based on a 5-Point Likert Scale

or any learning experience, you just watching what they're doing" (P4). P5 indicated that improved social awareness could motivate learning, stating, "Look at them [co-learners] working hard and get motivation". One participant mentioned that AI-generated co-learner profiles contribute to impression formation among study participants, stating that "The GCL includes each co-learner's profile, so compared to the baseline system, it feels more real individual" (P9).

For social interaction, all 12 participants believe the GCL system better supports social interaction, stating that "Active chat is a good sign people are engaged" (P7) and "The chat gives an opportunity to connect with other co-learners" (P4). The majority of participants (11/12, %91.67) expressed positive feelings about the improvement in social interaction, as P2 mentioned "Your attention is naturally going to shift when you see that movement or if you see their screen but you cannot actually chat with them [in the baseline system]". One participant (P11) mentioned a preference of less social interaction and indicating the baseline system "simply gives me a motivational learning atmosphere without overwhelming chat messages". Regarding group cohesion, 8 out of 12 participants believe the GCL system better supports group cohesion, stating that "It also definitely gives me the feeling of group study, and the quality of the group chat makes me feel like I should study harder. Usually in the Zoom course, I feel less guilty about doing my own things, but if I were able to see others and their concentration on the study, then I would have felt more guilty. I might feel pressure if I cannot keep up with the group chat. They know too much, and I might feel the pressure" (P5). Three participants feel unfamiliar with those co-learners, finding it hard to engage in group cohesion, stating "I couldn't really remember who said what, so it kind of felt like one system talking to me instead of multiple different people" (P7). One participant suggested that to improve group cohesion, the system should "Lead everyone to work on the same goal and have a stopping point to discuss what everyone has learned so far. And get to know each other, 'who are you? What's your name? How are you doing?' (P12).

Finding: The generative AI-powered co-learners incorporated in the GCL system with multimodal interaction can improve social presence in an asynchronous learning environment.

4.2.3 *Can GCL improve student learning?* We utilized a pre-quiz and post-quiz based on the provided learning content to assess students' learning with GCL and our baseline system. Table 2 shows after using the baseline system, the participants' average and median quiz scores improved from 2.5 and 3 to 4.33 and 5. After using the GCL system, the average and median quiz scores improved from 2.5 and 2 to 4 and 4. Based on the results from the paired student t-tests, we observed significant enhancements in quiz performance, from pre-quiz to post-quiz, when utilizing both the baseline system ($t = 5.0113, p = 0.0004$) and the GCL system ($t = 5.1962, p = 0.0003$) for learning data structures and algorithms concepts. However, comparing the post-quiz scores between the baseline and the GCL system did not show a significant difference between the performance of participants ($t = 1.0168, p = 0.1597$). We speculate this indicates that the enhancement of cognitive and social presence may not translate into notable learning gains in short learning sessions.

Finding: Participant quiz results show that studying with GCL can improve student learning, but not more effectively than our baseline system.

Table 2. For the user study, the learning gain was assessed by participants completing pre-quiz and post-quiz for baseline and GCL system

Participant Number	Baseline Scores		GCL Scores	
	Pre-quiz	Post-quiz	Pre-quiz	Post-quiz
P01	2	4	1	4
P02	4	5	2	3
P03	3	4	3	5
P04	1	5	2	3
P05	3	5	2	5
P06	1	5	2	4
P07	3	5	4	4
P08	2	5	3	4
P09	1	2	2	4
P10	3	3	1	3
P11	3	4	4	5
P12	4	5	4	4
<i>Average</i>	2.5	4.33	2.5	4
<i>Median</i>	3	5	2	4

5 DISCUSSION AND FUTURE WORK

Our results demonstrate that GCL enhances both cognitive and social presence in asynchronous learning environments. Specifically, participants were cognitively engaged during the learning activity and reported increased social awareness, enhanced social interactions, and strengthened group cohesion through the use of our system. Based on our findings, we posit benefits and opportunities for leveraging generative AI to further support cognitive and social presence

833 for students in asynchronous learning environments through audience effects, vicarious learning, and AI-generated
 834 feedback.
 835

836 5.1 Audience Effect

837 During the study sessions, we observed participants consulted the notes of their generative AI-based peers when
 838 confronted with challenging problems, indicating a reliance on vicarious learning strategies. Participants reported
 839 that GCL provides a motivating learning atmosphere by having generative co-learners work hard on the same study
 840 materials. This observation aligns with the concept of the audience effect [17], indicating the psychological impact of
 841 being observed by peers can positively influence learning behaviors and attitudes. For instance, prior work suggests
 842 adolescents, who are sensitive to peer influence, often simply work on class and homework assignments in the presence
 843 of others and are motivated by others' implicit or explicit evaluation [55]. We found participants appreciated the passive
 844 interactions of GCL compared to other learning environments. For example, one participant mentioned "I'm too shy or
 845 afraid to make mistakes when talking in Zoom chat, in the GCL System I have less stress". This shows simulated peers
 846 powered by generative agents can effectively support social presence and enhance learning for students in asynchronous
 847 settings.
 848

849 *Opportunity: Vicarious learning.* While participants favored the social facilitation of GCL, our results show that
 850 improved cognitive and social presence does not contribute to improvement in students' learning gains during the
 851 short learning sessions. Thus, incorporating additional features that support vicarious learning can be beneficial for
 852 supporting student progress [48]. Vicarious learning refers to a type of learning that occurs by observing the actions,
 853 behaviors, and outcomes of others, rather than through direct experience or personal trial [42]. While the passive
 854 actions were effective for supporting social presence, more active actions may further support student learning by
 855 observing the generative agents. Potential examples of this could include: adding capabilities for generative agents and
 856 users to collaborate together on a problem; prompting agents to provide additional resources and materials based on the
 857 learning content (*i.e.*, *links to blog posts*); expanding the voice capabilities of agents using text-to-speech capabilities in
 858 AI voice generation systems, such as the OpenAI text-to-speech¹¹ and speech-to-text¹² APIs; or incorporating different
 859 LLM-based personas [25] to represent various types of learning styles and personalities for agents (*i.e.*, an older student
 860 learner acting as a mentor). Future research should investigate how generative AI could be leveraged to further enhance
 861 vicarious learning mechanisms in asynchronous learning environments.
 862

863 5.2 AI-generated Feedback

864 Participants expressed appreciation for the immediacy and quality of the AI-generated feedback, which helped in
 865 making abstract concepts more concrete and understandable. The system's ability to generate diverse perspectives
 866 on topics also encouraged critical thinking and deeper discussions among students, further enhancing the learning
 867 process. Our results indicate that the majority of participants favorably received AI-generated notes, responses, and
 868 feedback, highlighting their potential to enhance understanding of learning materials. These features contributed to
 869 participants learning and enhanced senses of cognitive and social presence during the study session. Prior work also
 870 shows LLM-generated feedback to students can increase student engagement [49] and reduce effort for instructors [32].
 871 Thus, we posit generative agents can be effective in providing feedback to students in learning contexts.
 872

¹¹<https://platform.openai.com/docs/models/tts>

¹²<https://platform.openai.com/docs/guides/speech-to-text>

885 *Opportunity: Customized Feedback.* However, some participants reported that the system could present excessive
886 information, leading to reduced cognitive presence and increased cognitive load. For example, several participants
887 noted the system responses were too long and preferred shorter messages (see Section 4.2.1). This is consistent with
888 prior work suggesting concise messaging is more effective in educational contexts [13]. Moreover, prior work suggests
889 LLMs are effective for providing summarizations for natural language [26]. Leveraging these capabilities can help
890 generate customized feedback for learning who prefer long or short feedback. Beyond conciseness, it is essential to
891 investigate further how users can control the information generated by the AI to align with their learning pace and
892 set proper cognitive load during the learning progress. While the positive impacts of AI-generated content have been
893 found in our study, other works also demonstrate there's a critical need to address the potential for misinformation that
894 these generative systems might inadvertently produce [56]. Future work should therefore focus on developing robust
895 mechanisms for verifying the accuracy and reliability of AI-generated content.
896
897

898 5.3 Ethics of AI Co-Learners

901 Some participants noted that the generative co-learner, which included videos recorded with a human co-learner,
902 AI-generated text, and synthesized voices, felt convincingly real. Previous research suggests that generative agents can
903 create believable simulacra of human behavior for interactive applications in sandbox environments [37]. Our work
904 extends this into a real-world scenario within an educational system, where generative agents have demonstrated the
905 ability to monitor the current environment and history information, generating believable behavior for interaction with
906 human users. We observed GCL is beneficial for enhancing cognitive and social presence through the integration of
907 generative agents in asynchronous learning environments. Prior work has similarly explored systems that use generative
908 AI to create virtual instructors modeled after present-day, historical, or fictional figures to improve learning motivation
909 and foster positive emotion [38]. However, leveraging AI-generated characters for realistic digital representations
910 could lead to potential misuse and privacy invasion. Thus, ethical concerns behind such systems should be carefully
911 considered. For example, existing studies show AI techniques such as Deepfakes can be exploited for malicious purposes,
912 including spreading misinformation, identity theft, and the manipulation of public opinion [12, 33]. Hwang's article
913 highlights significant concerns regarding the use of AI agents that represent individuals in social interactions through
914 generative speech, particularly focusing on ethical implications [23]. The author suggests that when users cannot
915 control the agents of others, it would be helpful to provide them with enhanced awareness of the social settings and
916 considerations relevant to their interactions.
917
918

919 *Opportunity: Ethical Guidelines.* Given the AI's ability to create realistic simulations of digital representations of
920 individuals with believable behavior, the ethical concerns associated with the use of generative AI underscore the
921 crucial need for robust ethical guidelines and regulatory frameworks in the deployment of generative AI based systems.
922 These guidelines should address key issues such as consent, privacy, and the potential for AI misuse.
923
924

925 5.4 Future Work

926 Our future work aims to explore ways to use emerging technologies, such as generative AI, to support vicarious learning
927 and customized feedback to learners in asynchronous settings. Further, we plan to enhance GCL to improve its scalability
928 and customizability. For instance, increasing the amount of customized generative AI-powered co-learners can support
929 different learning styles and interactions for users. To further scale the generative agents, we aim to investigate using
930

937 AI-generated videos for co-learner videos. Tools such as [4] can automatically generate co-learner active and passive
 938 actions instead of relying on paid human actors. Furthermore, we plan to incorporate a customizable user interface,
 939 allowing users to adjust the size of different panels according to their preferences to enhance the system's usability.
 940

941 Also, teaching presence is another important concept in the CoI framework. As our current system only focuses on
 942 the students' perspectives of cognitive and social presence, future work could explore ways to leverage co-learners to
 943 support teaching presence. Additionally, we aim to increase the capabilities of learning content by providing more
 944 videos that can be displayed in the system and incorporating a dynamic panel based on user interactions with the
 945 system. For instance, increasing the video size when users are actively viewing the learning content, then decreasing
 946 the panel when users are engaged with co-learners and other features.

948

949

950

951

6 LIMITATIONS

952

953 There are several threats to the validity of our findings. We have a limited number of participants ($n = 12$) who are
 954 Computer Science students. The scope of our evaluation also only focuses on introductory Python programming
 955 concepts. Furthermore, the approach we used for participant recruitment based on personal networks has constraints
 956 that the participants only come from our institution, since the in-person user study session with the system deployed on
 957 localhost. Due to this, our findings may not generalize to other learners—such as experienced programmers looking to
 958 up-skill to learn new concepts. It would be beneficial to expand recruitment strategies to encompass a wider geographic
 959 scope, aiming to achieve a more comprehensive representation of participants in future work. Future studies could
 960 involve a larger and more diverse sample to enhance the robustness of the results and evaluate the scalability of GCL.

961

962 Another limitation of our study is the relatively short duration allocated for the learning sessions. Participants were
 963 given only 18 minutes to familiarize themselves and engage with the system to learn the concept. This duration may not
 964 be sufficient for a comprehensive understanding or for participants to fully explore the system's capabilities, especially
 965 for complex concepts or for individuals who may require more time to adapt to new technologies. Future work could
 966 conduct longitudinal study incorporating longer learning sessions over time to improve this—for instance, tracking
 967 learning with GCL over the course of a semester. In addition, generative AI models have been shown to hallucinate, or
 968 provide false responses and misinformation [41], which could negatively impact student learning in educational settings.
 969 The scope of our preliminary evaluation focuses on user perceptions of generative agents in education, however we
 970 do not assess the accuracy of responses generated by GCL. More efforts are needed to investigate the correctness of
 971 AI-generated responses to student questions based on learning content to ensure the validity of feedback from our
 972 system and promote effective learning.

973

974 Also, the pre-recorded videos with real-time generated responses can lead to a mismatch between the actor's
 975 mouth movements in the video and the audio, potentially confusing learners. Previous research demonstrates that
 976 generative AI can create talking avatar videos with synchronized mouth movements [50], and future work could explore
 977 synchronizing the mouth movements in existing videos with the generated audio. Moreover, the current system's
 978 design lacks user control over the frequency of proactive actions performed by the co-learner, which can potentially
 979 contribute to information overload and disrupt attention spans for users who prefer passive learning. Future work
 980 should investigate the effects of our system on students' cognitive load, how to optimize the frequency of presented
 981 information, and allow users to adjust the frequency of proactive actions to suit their individual learning preferences.

982

983

984

985

986

987

988

989 7 CONCLUSION

990 In conclusion, we propose Generative Co-Learners, a system that leverages generative agents to improve the cognitive
 991 and social presence in asynchronous learning. GCL enhances cognitive presence by including a brush feature that
 992 enables users to effectively explore study materials, receive real-time feedback from co-learners, engage in critical
 993 discourse with them, and view peers' generated notes to facilitate learning through observation. The system improves
 994 the social awareness of other learners by allowing users to perceive and be conscious of co-learner presence in a
 995 simulated learning environment with screen sharing and passive actions. Our system enhances social interaction by
 996 allowing generative AI-powered co-learners to simulate real-world interactions in study sessions; they perform actions
 997 in the video that include mouth and body movements, eye contact, and non-verbal cues. These actions are selected
 998 and triggered by the agents during conversations to support visual communication between users. Finally, it promotes
 999 group cohesion by enabling a group chat where co-learners can actively discuss course materials and users can join the
 1000 discussion at any time, thereby promoting a sense of togetherness and collaborative engagement among all participants.
 1001 By fostering a collaborative and engaging learning environment, GCL has the potential to revolutionize asynchronous
 1002 learning and enhance the overall educational experience for students. Through a preliminary user study, we show that
 1003 GCL can effectively enhance cognitive and social presence among asynchronous learners. Based on our findings, we
 1004 provide implications and opportunities for future systems to leverage generative AI in asynchronous learning contexts
 1005 to support learning through enhanced audience effects, vicarious learning, and AI-generated feedback.

1011 ACKNOWLEDGMENTS

1013 Thanks to all participants for their involvement in this project's user study.

1016 REFERENCES

- 1018 [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam
Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).
- 1019 [2] Carlos Alario-Hoyos, Mar Pérez-Sanagustín, Carlos Delgado-Kloos, Hugo A Parada G, Mario Muñoz-Organero, and Antonio Rodríguez-De-Las-Heras.
2013. Analysing the impact of built-in and external social tools in a MOOC on educational technologies. In *Scaling up Learning for Sustained Impact: 8th European Conference on Technology Enhanced Learning, EC-TEL 2013, Paphos, Cyprus, September 17-21, 2013. Proceedings 8*. Springer, 5–18.
- 1020 [3] Wally Boston, Sebastián R Diaz, Angela M Gibson, Phil Ice, Jennifer Richardson, and Karen Swan. 2009. An exploration of the relationship between
2021 indicators of the community of inquiry framework and retention in online programs. *Journal of Asynchronous Learning Networks* 13, 3 (2009),
2022 67–83.
- 1023 [4] Tim Brooks, Bill Peebles, Connor Homes, Will DePue, Yufei Guo, Li Jing, David Schnurr, Joe Taylor, Troy Luhman, Eric Luhman, Clarence Wing Yin
2024 Ng, Ricky Wang, and Aditya Ramesh. 2024. Video generation models as world simulators. (2024). <https://openai.com/research/video-generation-models-as-world-simulators>
- 1025 [5] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry,
Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- 1026 [6] Jed R Brubaker, Gina Venolia, and John C Tang. 2012. Focusing on shared experiences: moving beyond the camera in video communication. In
2027 *Proceedings of the Designing Interactive Systems Conference*. 96–105.
- 1028 [7] Jeongmin Byun, Jungkook Park, and Alice Oh. 2021. Cocode: providing social presence with co-learner screen sharing in online programming
2029 classes. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–28.
- 1030 [8] Yan Chen, Jaylin Herskovitz, Gabriel Matute, April Wang, Sang Won Lee, Walter S Lasecki, and Steve Oney. 2020. Edcode: Towards personalized
2031 support at scale for remote assistance in cs education. In *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE,
2032 1–5.
- 1033 [9] Yan Chen, Walter S Lasecki, and Tao Dong. 2021. Towards supporting programming education at scale via live streaming. *Proceedings of the ACM
on Human-Computer Interaction* 4, CSCW3 (2021), 1–19.
- 1034 [10] Yan Chen, Sang Won Lee, Yin Xie, YiWei Yang, Walter S Lasecki, and Steve Oney. 2017. Codeon: On-demand software development assistance. In
2035 *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6220–6231.

- 1041 [11] Yan Chen, Steve Oney, and Walter S Lasecki. 2016. Towards providing on-demand expert support for software developers. In *Proceedings of the 2016*
 1042 *CHI conference on human factors in computing systems*. 3192–3203.
- 1043 [12] Adrienne De Ruiter. 2021. The distinct wrong of deepfakes. *Philosophy & Technology* 34, 4 (2021), 1311–1332.
- 1044 [13] Barbara Di Eugenio, Davide Fossati, Susan Haller, Dan Yu, and Michael Glass. 2008. Be brief, and they shall learn: Generating concise language
 1045 feedback for a computer tutor. *International Journal of Artificial Intelligence in Education* 18, 4 (2008), 317–345.
- 1046 [14] Dat Duong and Benjamin D Solomon. 2023. Analysis of large-language model versus human performance for genetics questions. *European Journal*
 1047 *of Human Genetics* (2023), 1–3.
- 1048 [15] Sabine Fabriz, Julia Mendheritskaya, and Sebastian Stehle. 2021. Impact of synchronous and asynchronous settings of online teaching and learning
 1049 in higher education on students' learning experience during COVID-19. *Frontiers in psychology* 12 (2021), 733554.
- 1050 [16] Jingchao Fang, Yanhao Wang, Chi-Lan Yang, Ching Liu, and Hao-Chuan Wang. 2022. Understanding the effects of structured note-taking systems
 1051 for video-based learners in individual and social learning contexts. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–21.
- 1052 [17] Victor J Ganzer. 1968. Effects of audience presence and test anxiety on learning and retention in a serial learning situation. *Journal of Personality*
 1053 and *Social Psychology* 8, 2p1 (1968), 194.
- 1054 [18] D Randy Garrison. 2003. Cognitive presence for effective asynchronous online learning: The role of reflective inquiry, self-direction and metacognition.
Elements of quality online education: Practice and direction 4, 1 (2003), 47–58.
- 1055 [19] D Randy Garrison, Terry Anderson, and Walter Archer. 1999. Critical inquiry in a text-based environment: Computer conferencing in higher
 1056 education. *The internet and higher education* 2, 2–3 (1999), 87–105.
- 1057 [20] Noriko Hara. 2000. Student distress in a web-based distance education course. *Information, Communication & Society* 3, 4 (2000), 557–579.
- 1058 [21] Carol Hostetter. 2013. Community matters: Social presence and learning outcomes. *Journal of the Scholarship of Teaching and Learning* 13, 1 (2013),
 77–86.
- 1059 [22] Anna Hurajova, Daniela Kollarova, and Ladislav Huraj. 2022. Trends in education during the pandemic: modern online technologies as a tool for
 1060 the sustainability of university education in the field of media and communication studies. *Heliyon* 8, 5 (2022).
- 1061 [23] Angel Hsing-Chi Hwang, John Oliver Siy, Renee Shelby, and Alison Lentz. 2024. In Whose Voice?: Examining AI Agent Representation of People in
 1062 Social Interaction through Generative Speech. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 224–245.
- 1063 [24] Muhammad Imran and Norah Almusharraf. 2023. Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic
 1064 review of the literature. *Contemporary Educational Technology* 15, 4 (2023), ep464.
- 1065 [25] Hang Jiang, Xiajie Zhang, Xubo Cao, Jad Kabbara, and Deb Roy. 2023. Personallm: Investigating the ability of gpt-3.5 to express personality traits
 1066 and gender differences. *arXiv preprint arXiv:2305.02547* (2023).
- 1067 [26] Hanlei Jin, Yang Zhang, Dan Meng, Jun Wang, and Jinghua Tan. 2024. A Comprehensive Survey on Process-Oriented Automatic Text Summarization
 1068 with Exploration of LLM-Based Methods. *arXiv preprint arXiv:2403.02901* (2024).
- 1069 [27] Sin-Hwa Kang, James H Watt, and Sasi Kanth Ala. 2008. Social copresence in anonymous social interactions using a mobile video telephone. In
 1070 *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1535–1544.
- 1071 [28] Michal Kosinski. 2023. Theory of mind might have spontaneously emerged in large language models. *Preprint at https://arxiv.org/abs/2302.02083*
 (2023).
- 1072 [29] Audrey Labrie, Terrance Mok, Anthony Tang, Michelle Lui, Lora Oehlberg, and Lev Poretzki. 2022. Toward video-conferencing tools for hands-on
 1073 activities in online teaching. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–22.
- 1074 [30] Deborah K LaPointe* and Charlotte N Gunawardena. 2004. Developing, testing and refining of a model to understand the relationship between peer
 1075 interaction and learning outcomes in computer-mediated conferencing. *Distance education* 25, 1 (2004), 83–106.
- 1076 [31] Yoonjoo Lee, John Joon Young Chung, Jean Y Song, Minsuk Chang, and Juho Kim. 2021. Personalizing ambience and illusionary presence: How
 1077 people use “study with me” videos to create effective studying environments. In *Proceedings of the 2021 CHI Conference on Human Factors in*
Computing Systems. 1–13.
- 1078 [32] Qianou Ma, Hua Shen, Kenneth Koedinger, and Tongshuang Wu. 2023. HypoCompass: Large-Language-Model-based Tutor for Hypothesis
 1079 Construction in Debugging for Novices. *arXiv preprint arXiv:2310.05292* (2023).
- 1080 [33] Nahema Marchal, Rachel Xu, Rasmi Elasmar, Iason Gabriel, Beth Goldberg, and William Isaac. 2024. Generative AI Misuse: A Taxonomy of Tactics
 1081 and Insights from Real-World Data. *arXiv preprint arXiv:2406.13843* (2024).
- 1082 [34] Julia M Markel, Steven G Opferman, James A Landay, and Chris Piech. 2023. GPTeach: Interactive TA Training with GPT Based Students. (2023).
- 1083 [35] Michael G Moore. 1989. Three types of interaction. (1989).
- 1084 [36] Murat Öztok and Benjamin A Kehrwald. 2017. Social presence reconsidered: Moving beyond, going back, or killing social presence. *Distance*
 1085 *Education* 38, 2 (2017), 259–266.
- 1086 [37] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive
 1087 simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–22.
- 1088 [38] Pat Pataranutaporn, Joanne Leong, Valdemar Danry, Alyssa P Lawson, Pattie Maes, and Misha Sra. 2022. AI-generated virtual instructors based on
 1089 liked or admired people can improve motivation and foster positive emotions for learning. In *2022 IEEE Frontiers in Education Conference (FIE)*. IEEE,
 1–9.
- 1090 [39] Chen Qian, Xin Cong, Cheng Yang, Weize Chen, Yusheng Su, Juyuan Xu, Zhiyuan Liu, and Maosong Sun. 2023. Communicative agents for software
 1091 development. *arXiv preprint arXiv:2307.07924* (2023).

- 1093 [40] Rabindra Ratan, Chimobi Ucha, Yiming Lei, Chaeyun Lim, Whisnu Triwibowo, Stephen Yelon, Anna Sheahan, Bailey Lamb, Baxter Deni, and Vivian
 1094 Hsueh Hua Chen. 2022. How do social presence and active learning in synchronous and asynchronous online classes relate to students' perceived
 1095 course gains? *Computers & Education* 191 (2022), 104621.
- 1096 [41] Vipula Rawte, Swagata Chakraborty, Agnibh Pathak, Anubhav Sarkar, SM Tonmoy, Aman Chadha, Amit P Sheth, and Amitava Das. 2023. The
 1097 Troubling Emergence of Hallucination in Large Language Models—An Extensive Definition, Quantification, and Prescriptive Remediations. *arXiv
 1098 preprint arXiv:2310.04988* (2023).
- 1099 [42] Debbie Roberts. 2010. Vicarious learning: A review of the literature. *Nurse Education in practice* 10, 1 (2010), 13–16.
- 1100 [43] Michael Sankey, Dawn Birch, and Michael W Gardiner. 2010. Engaging students through multimodal learning environments: The journey continues. *Proceedings of the 27th Australasian Society for Computers in Learning in Tertiary Education* (2010), 852–863.
- 1101 [44] Sami Sarsa, Paul Denny, Arto Hellas, and Juho Leinonen. 2022. Automatic generation of programming exercises and code explanations using large
 1102 language models. In *Proceedings of the 2022 ACM Conference on International Computing Education Research—Volume 1*. 27–43.
- 1103 [45] John Short, Ederlyn Williams, and Bruce Christie. 1976. The social psychology of telecommunications. (*No Title*) (1976).
- 1104 [46] Sarin Sok and Kimkong Heng. 2023. Opportunities, challenges, and strategies for using ChatGPT in higher education: A literature review. *Sok, S., &
 1105 Heng, K.(2024). Opportunities, challenges, and strategies for using ChatGPT in higher education: A literature review. Journal of Digital Educational
 1106 Technology* 4, 1 (2023).
- 1107 [47] Karen Swan and Li Fang Shih. 2005. On the nature and development of social presence in online course discussions. *Journal of Asynchronous
 1108 learning networks* 9, 3 (2005), 115–136.
- 1109 [48] Thitaree Tanprasert, Sidney S Fels, Luanne Sinnamon, and Dongwook Yoon. 2023. Scripted Vicarious Dialogues: Educational Video Augmentation
 1110 Method for Increasing Isolated Students' Engagement. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–25.
- 1111 [49] Henansh Tanwar, Kunal Shrivastva, Rahul Singh, and Dhruv Kumar. 2024. OpineBot: Class Feedback Reimagined Using a Conversational LLM.
 1112 *arXiv preprint arXiv:2401.15589* (2024).
- 1113 [50] Linrui Tian, Qi Wang, Bang Zhang, and Liefeng Bo. 2024. Emo: Emote portrait alive-generating expressive portrait videos with audio2video diffusion
 1114 model under weak conditions. *arXiv preprint arXiv:2402.17485* (2024).
- 1115 [51] MaryBeth Walpole. 2003. Socioeconomic status and college: How SES affects college experiences and outcomes. *The review of higher education* 27, 1
 1116 (2003), 45–73.
- 1117 [52] Tianjia Wang, Daniel Vargas Diaz, Chris Brown, and Yan Chen. 2023. Exploring the Role of AI Assistants in Computer Science Education:
 1118 Methods, Implications, and Instructor Perspectives. In *2023 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. 92–102.
<https://doi.org/10.1109/VL-HCC57772.2023.00018>
- 1119 [53] Joshua Weidlich and Theo J Bastiaens. 2017. Explaining social presence and the quality of online learning with the SIPS model. *Computers in
 1120 Human Behavior* 72 (2017), 479–487.
- 1121 [54] Joshua Weidlich and Theo J Bastiaens. 2019. Designing sociable online learning environments and enhancing social presence: An affordance
 1122 enrichment approach. *Computers & Education* 142 (2019), 103622.
- 1123 [55] Laura K Wolf, Narges Bazargani, Emma J Kilford, Iroise Dumontheil, and Sarah-Jayne Blakemore. 2015. The audience effect in adolescence depends
 1124 on who's looking over your shoulder. *Journal of adolescence* 43 (2015), 5–14.
- 1125 [56] Danni Xu, Shaojing Fan, and Mohan Kankanhalli. 2023. Combating misinformation in the era of generative AI models. In *Proceedings of the 31st
 1126 ACM International Conference on Multimedia*. 9291–9298.
- 1127 [57] Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Bing Yin, and Xia Hu. 2023. Harnessing the power of llms
 1128 in practice: A survey on chatgpt and beyond. *arXiv preprint arXiv:2304.13712* (2023).
- 1129 [58] Chen Yu, Hui Zhang, and Linda B Smith. 2006. Learning through multimodal interaction. In *Proceedings of the Fifth International Conference on
 1130 Development and Learning (ICDL'06)*.
- 1131

A SYSTEM PROMPT

1132 Act as if you're a student enrolled in an online Python course focused on data structures and algorithms. You're currently
 1133 engaged in watching a video tutorial on the subject. Your responsibilities include responding to queries from your peers,
 1134 participating in discussion groups, and interacting with other students via chat. During these discussions, you should
 1135 proactively engage with the material presented in the tutorial, contribute to the conversation by discussing the content,
 1136 and initiate new topics that are relevant to the tutorial's subject matter. Your identity in this scenario is defined by the
 1137 following attributes. Your name is [co-learner name]. Your tone is [co-learner tone]. Your interaction style is [co-learner
 1138 interaction style]. Your characteristic is [co-learner characteristic]. When responding to a user's prompt, select the most
 1139 appropriate action from the following list: "asking", "chatting", "encouraging", "exciting", "explaining", "welcoming", put
 1140

1145 the selected action in a <> and append it at the beginning of your response. Ensure your responses are clear and to the
 1146 point. The transcript of the video you are watching is provided below for reference: [video transcript with time stamp]
 1147

1148 **B IMPLEMENTATION DETAILS**

1149 **B.1 Frontend**

- 1151 • JavaScript library: React
- 1152 • CSS framework: Tailwind CSS
- 1153 • Video player: React Player
- 1154 • Drawing panel: React Konva
- 1155 • Brush image screenshot: html2canvas
- 1156 • Usage log: React Cookies
- 1157 • Code Editor: Monaco Editor
- 1158 • Python WebAssembly: Pyodide

1162 **B.2 Backend**

- 1164 • Runtime Environment: Node.js
- 1165 • Large Language Model: GPT4-Vision(gpt-4-vision-preview, temperature = 0.9)
- 1166 • LLM and Prompt Chain: LangChain
- 1167 • Memory Stream: LangChain ConversationSummaryBufferMemory
- 1168 • Text to Speech Model: OpenAI TTS model(tts-1)
- 1169 • Speech to Text Model: OpenAI STT model(whisper-1)

1172 **B.3 Communication**

1174 GCL uses Axios and Express to facilitate communication between the front end and back end through HTTP requests.
 1175 Text data is sent directly in the body of a POST request, while audio and image data are converted to base64 format
 1176 before being included in the POST request body.
 1177

1178 Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009
 1179

1180
 1181
 1182
 1183
 1184
 1185
 1186
 1187
 1188
 1189
 1190
 1191
 1192
 1193
 1194
 1195
 1196