Journal of Sociocybernetics

Journal of Sociocybernetics ISSN 1607-8667

ORIGINAL ARTICLE

AI as a Socratic Opponent: Comparative network analyses of class discussions from a college psychology course

Shantanu Tilak, Baylee Brown, Hovhannes Madanyan, Gabriella Washington, Jazzmin Collier, Rebecca Ragnedda, Jaiden Mitchell, Jasha Brewington, Briana Hall, Kristopher Barnum, Courtney Moore, Allure Harris, Beckham Rombaoa, Jayden Evans, Emily Shipp, Makenzie Short, Jamal Thomas, Carmello Browne, Hassan Abbasi, Trent Hammer, Nathan C. Prince, Kadie F. Kennedy

Virginia Wesleyan University, Virginia Beach, VA, United States of America

This mixed methods participatory study was co-authored by 19 undergraduate students and their instructor in an introductory psychology class, with help from two research assistants. Participant observers evaluated and reflected upon the use of artificial intelligence (AI) language models as surrogate agents to support classroom discussion forums. An M- and P-individual framework rooted in Gordon Pask's cybernetics is used to structure out human-computer interaction feedback loops that ensued during class discussions. Live chats were held during each lecture on a Google community, wherein students would respond to a weekly prompt posted by the instructor and to peers. Two of these sessions were held on the Character.AI and DeepAI platforms. Four groups of students interacted with language models of Freud and Piaget during sessions related to human consciousness and development, with one student "driver" prompting the AI after group brainstorming. Temporally proximal business-as-usual chats on the nervous system and human learning are compared to AI discussions using the igraph network analysis package in RStudio. Comparative network visualizations highlight the possibility to create decentralized discussions using AI in college classrooms. To better understand student-to-student interactions guiding the driver's prompting in AI chats, qualitative insights are shared from each group.

10.26754/ojs_jos/jos.2025111783

Journal of Sociocybernetics 20(1) (2025)

Correction (26th August 2025): *M- and P- blueprints were referred to in a previous version of this paper as embodiment and cyclicity diagrams. This error has been corrected, as cyclicity is only observed in entailment meshes where concepts can be derived from one another.*

1. Introduction

The trajectory of the development of artificial intelligence (AI) can be compared to the roadmap taken in diversifying the Internet (Sharples, 2023). The web, which began as a technology for retrieving data/information, evolved to support social interactions between users. Similarly, AI was initially designed to solve narrow problems, beginning as a project in the 1950's posts the Dartmouth Conference under the premise that:

"Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." (McCarthy et al., 1955, p.12)

Since then, AI has been scaled to respond to human input using multimedia output. The evolution of both technologies began with a single corporation. The European Organization for Nuclear Research (CERN) was the progenitor of the web, and OpenAI of AI, respectively.

The ultimate promise of AI is the idea that disruptive advances could enable AI to reach/exceed human-level intelligence (Bostrom, 2014). After Dartmouth, research related to AI split into three schools. McCarthy focused on formal logic. Minsky wanted to create AI as a product of brute engineering transcending logic. Herbert Simon focused on conceptualizing AI as displaying intelligence through its capacity for problem-solving and decision-making, inspired by cognitive psychology (Halpin, 2025).

The vision of AI as an autonomous agent has obscured possibilities associated with human agency in using it. Particularly, the potentials of a distributed paradigm of AI, treating human-computer interaction as emergent feedback loops between artificial and living agents have not been fully unearthed (Sharples, 2023). Instead, a focus has been laid on individualistic understandings of human-AI interaction. Interest in collaboration and conversational landscapes, associated with research in cybernetics and constructivism dwindled in academic programs focused on AI (Tilak et al., 2022).

In this mixed methods action research study, we expand upon extant theoretical literature focused on using AI in collaborative learning (Sharples, 2023). We showcase one role that could be assigned to AI in collaborative learning through the example of an introductory college psychology class. Groups of students led by a representative learner and an instructor interacted with AI chatbots of eminent figures in psychology (namely, Freud and Piaget). Our theoretical framework sets up the approach taken in our study, outlining roles AI can play in collaborative classrooms, reviewing extant literature on its educational use, and visualizing activities to be conducted in the current study using basic cybernetic principles.

2. The collaborative possibilities of Artificial Intelligence

Contemporary AI tools like language models are often studied using a series of one-to-one human-computer prompts (Sabzalieva & Valentini, 2023). This paradigm can be compared to stimulus-response mechanisms in behaviorist psychology; sometimes too simplistic an approach to understand how and why human beings learn and adapt. If tools are to augment human cognition and behavior and simulate human functioning, they must be used in contexts that tap into the social abilities of humans (Tilak et al., 2022). Accordingly, AI tools can be treated as artificial component in networks of dynamic living minds. This possibility is intuitive in a context where AI has been integrated into social media and workplace technology suites. A collaborative paradigm of artificial intelligence is not new, just underexplored; it was foreseen by cybernetician Gordon Pask (1975,76) as early as the 60s.

Cybernetics is a transdiscipline that investigates how living and artificial systems (machines, cells, organs, brains, humans) interact and respond to their environments in real-time (Tilak et al., 2022). A human-AI problem-solving scenario can be designed as a distributed landscape using cybernetic principles. Contemporary scholars suggest systems of such activity are sociotechnical since interactions occurring in them are supported by human-computer feedback loops devoted towards targeted problem-solving (Behymer & Flach, 2016). Designers can understand the nature of interactions in these systems by collaborating with human agents participating in them to hear their insights. From participatory action research (PAR), one can infer how systems can be set up to be productive and meaningful using data-driven insights, and how they can be redesigned (Glassman et al., 2013). Observation while being a member of the system involved in the learning process is a particularly powerful method in PAR.

Nimbly implementing participatory research in real-time at the confluence of education, computer science, psychology, and design was the broader goal of Pask's work (De Zeeuw, 2001). Pask's focus on dialogic learning, collaboration, and instructional design is emblematic of the famous developmental psychologist Lev Vygotsky's work (see Pask, 1966, p.226; 1976, p.19-20, and Tilak & Glassman, 2022). His definition of cybernetics (Tzafestas et al., 2017), presented below, highlights this connection:

"Cybernetics is the science or the art of manipulating defensible metaphors, showing how they may be constructed and what can be informed as a result of their existence." (p.124)

However, Pask's foretelling of AI, and focus on adaptive technologies highlights how his work acts as a refashioning of cultural historical psychology for the Information Age. While contemporary Vygotskian scholars (Fleer, 2016; Rubtsova & Salomatova, 2022) have formulated theories of children's digital play focusing on student activity with screens and tools to fit Vygotsky's theory to current societal conditions, Pask's approach can add a rigorous human-centered design dimension to educational technology research, and even user experience (UX) design research at large.

In human-AI sociotechnical systems, AI can become a surrogate agent (Scott, 2016) that responds to human participants as collaborative activities ensue. Humans must become dominant problem-solvers and prompt AI to support ongoing activity rather than treat it as a "crutch" that performs tasks for them. Consonant with Licklider's (1960) early ideas about symbiosis between humans and technology akin to the fig tree and wasp, AI should augment human thinking and action, not substitute it.

Sharples (2023), and Sabzalieva & Valentini (2023), who are jointly attributed in a UNESCO publication, outline six roles AI can adopt in collaborative learning to assist working groups of students:

1. **Possibility Engine:** Generates alternative ways to look at information.
2. **Socratic Opponent:** Helps consider alternate perspectives in a discussion.
3. **Collaborative Coach:** Assists/guides joint problem-solving.
4. **Co-Designer:** Helps create artifacts to present usable information.
5. **Exploratorium**: Helps understand ways to present, analyze and develop insights from data.
6. **Storyteller:** Helps tell a story narrating the variety of experiences that can be undertaken in our world.

Previous work conducted in our institution's research program, run as a partnership between a liberal arts university and special education school has responded to Sharples' call for action. We have investigated the possibility for teachers to use AI language models as an assistant in curriculum blueprint creation (as a possibility engine and co-designer; Tilak et al., 2024a), their use by college students in generating multimodal classroom artifacts (as a storyteller, possibility engine, and co-designer; Tilak et al., 2024b), and the use of autonomous learning technologies to explore student problem-solving data to improve K-12 curricula in real-time (as a data exploratorium and co-designer; Tilak & Bogacki, 2024). These studies were published in 2024, in the *Journal of Sociocybernetics* and presented at the *American Society for Cybernetics 60th Anniversary Meeting*. Here, we continue our efforts to give life to Sharples' (2023) ideas, exploring AI's specific use as a Socratic Opponent in classroom discussions.

3. AI as a Socratic Opponent: Extant Literature

In this section, we review a selection of studies focusing on AI's use in shepherding conversation/argumentation. Language models can be embedded into classrooms as a tool that facilitates Socratic questioning, to encourage exploration of ideas and thinking about alternative arguments. High schoolers have been shown to be able to socially engage one-on-one with chatbots that ask questions step-by-step and display more behavioral engagement, or conversational turn-taking (Blasco & Charisi, 2024). However, when it comes to using AI to help explore concepts, there are limitations associated with using it as the initiator of questioning. Students suggest that AI's line of questioning can be tangential.

In college settings, inquiry into using ChatGPT as a Socratic questioning tool has shown that while students appreciate the cold, hard facts AI can provide, instructor opinions/feedback related to classroom discussions are hard to replace in courses relying on argumentation and perspective taking (Fakour & Imani, 2025). In therapy contexts, AI models have been used with a co-peer-based system to rate/supervise responses before they are presented to a client to address mental health concerns. However, even a small chance for harmful responses/questions presents immense liability concerns (Held et al., 2024).

While studies mentioned above lay focus on Socratic questioning in one-to-one human-AI systems, group dynamics in prompting and responding to AI models have also been studied. Haqbeen et al. (2023) investigated the d-agree AI discussion tool's utility to facilitate assertive civic discussion online. AI-facilitated chats led to greater like button use and replies, indicating higher social participation. Kim et al. (2020), in their sample of 134 adult participants, saw that GroupfeedBot was able to increase egalitarian talk in medium sized groups of around 25 individuals, and encourage more diverse opinions (measured by gauging the unique morphemes in business-as-usual and chatbot mediated discussions) about topics being discussed. In Do et al.'s (2022) study with 42 adult groups in online settings, a group chat-based bot that would encourage participation from socially isolated individuals was gauged in terms of efficacy in promoting egalitarian conversation. Those that were given false negative prompts about lack of participation accepted the use of AI to a greater degree, while those that were prompted wrongly (false positive) to participate more showed greater social engagement. While these studies highlight the possibility to use AI to mediate a group discussion, the opinions and views of human agents about AI's role, and conversational network characteristics are only explored in a cursory manner.

The described studies have mostly showcased how AI can be used to facilitate and prompt more questioning/exploration rather than letting human agents take on this role to ask for objective facts from AI they can vet and critique. Secondly, the use of AI to facilitate group discussions and comparative analyses of AI and non-AI discussion settings has mainly been accompanied by static frequencies of like, comment, word count and conversational turns. The use of network statistics (Kolaczyk & Csárdi, 2014) that understand the evolution of a conversational system as a dynamic group of feedback loops can help understanding emergent, reflexive processes at play. Moreover, sharing the perspectives of living agents in these conversations about the limitations and victories of using AI can add an experiential dimension to current research efforts.

To expand current efforts, we use a framework adopting Gordon Pask's (1975) M- and P-individual nomenclature to understand mechanisms at play in human-AI conversations. In our study, we compare AI-mediated and business-as-usual classroom discussions.

4. Conversation Theory: Visualizing human-computer interaction

Gordon Pask developed two cybernetic approaches to expose the sociocognitive mechanisms of machine-dependent and independent conversations between humans and/or artificial systems. The first was called conversation theory (CT) and is used to decode mechanisms at play in strict conversational environments guided by specific topics, such as classrooms. The second, interaction of actors theory (IA, a play on AI), focuses on potentially endless everyday conversations not bounded by a specific topic or focus (De Zeeuw, 2001), such as texts between friends.

Both approaches utilize an analytic distinction, suggesting that all mechanical bodies (brains, computers) with physical presence process ideas and perceptions in context, akin to hardware and software (Pask, 1975). Materially "present" systems are M-individuals or mechanical individuals. Concepts processed at the boundaries between M-individuals are P-individuals, or psychological individuals. In essence, these systems create networks that are organizationally closed, and informationally open, since M-individuals can embody one or more emergent P-individuals (i.e., a P-individual is always embodied in one or more M-individuals). P-individuals are interrelated to one another in what are known as entailment meshes (Glanville & Pak, 2010).

Entailment meshes can be used to program technologies and power them with information. However, a simpler M- and P- individual framework can find use to create design blueprints of proposed and/or observed human-computer interaction in any context (Tilak et al., 2024c). Analyzing the nature of human-computer conversations by recounting words/ideas exchanged, or measuring the strength of distributed activity, one can understand the efficacy of a system or activity configuration designed for collaborative learning. Data analysis can be utilized to make inferences to refine the system and shepherd productivity. Conversations aid design, and design aids conversation (Pangaro, 2008).

The end-goal (broadly) of using Paskian cybernetics would be to better meet the needs of human participants to effectively engage in joint activity. In essence, both CT and IA can add rigor to the facilitation and improvement of collaborative learning contexts. Both sharp and fuzzy methods relying on numerical and string data can support designing and (re)designing such environments (Westermann, 2018). Users and designers should become participant observers in these conversational landscapes, especially in the context of studies framed within second-order cybernetics.

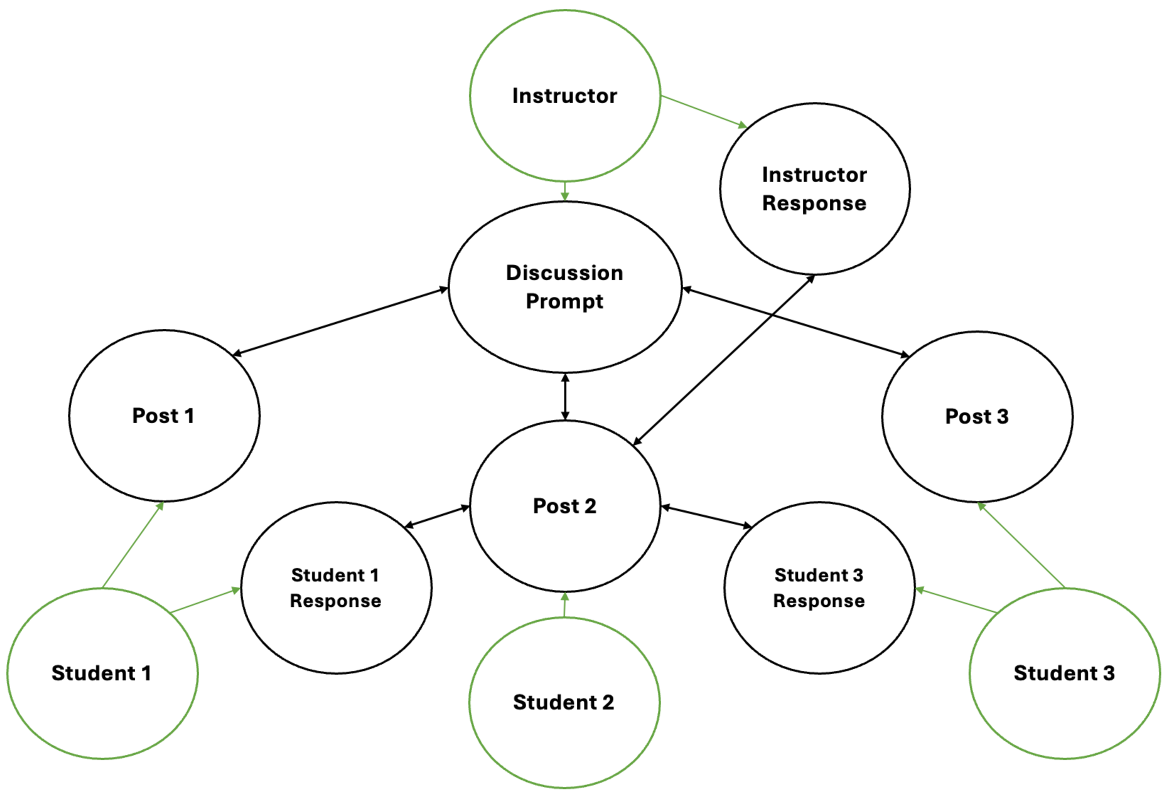
Our work blurs the boundary between the researcher and participants in formulating how to lay out human-computer interaction in a classroom. Rather than taking on an ethnomethodological approach to have external observers enter the classroom, prescribe study methods, observe how classroom communities evolve, or even record video data (e.g., Haataja et al., 2022), we use an intramundane source of truth to maintain the naturalistic state of the classroom and understand the firsthand experiences of participant observers. The idea of a Hawthorne effect (Oswald et al., 2014) that influences participant activity in cases of awareness of external observation informs our decision to utilize a radical participatory methodology. Students and their instructor in this study function as co-designers applying a second-order cybernetic pedagogy (Reinertsen, 2012), collecting data from AI-mediated classroom activities, and reflecting on these interactions. Quantitative network analysis that captures emergent conversational mechanisms at play between living and AI systems (Kolaczyk & Csárdi, 2014) is used to interpret classroom data, along with qualitative narrative reflection (sharp and fuzzy methods; Westermann, 2018).

We used Pask's nomenclature to create preliminary blueprints of interactions in our classroom, for both business-as-usual discussions on Google chat, and AI-mediated chats conducted with chatbots on the Character.AI and DeepAI platforms. Two M- and P- blueprints were crafted as part of research procedures ensuing during our participatory classroom project. Both followed the most rudimentary principles of CT. The first blueprint was a process diagram of business-as-usual classroom discussions. The instructor would post a weekly prompt to the Google Chat Community, and students would respond to it and (sometimes) each other.

An example with three students responding to the instructor's prompt and to each other is shown in Figure 1.

Figure 1

Blueprint of business-as-usual (non-AI) chat flowchart.

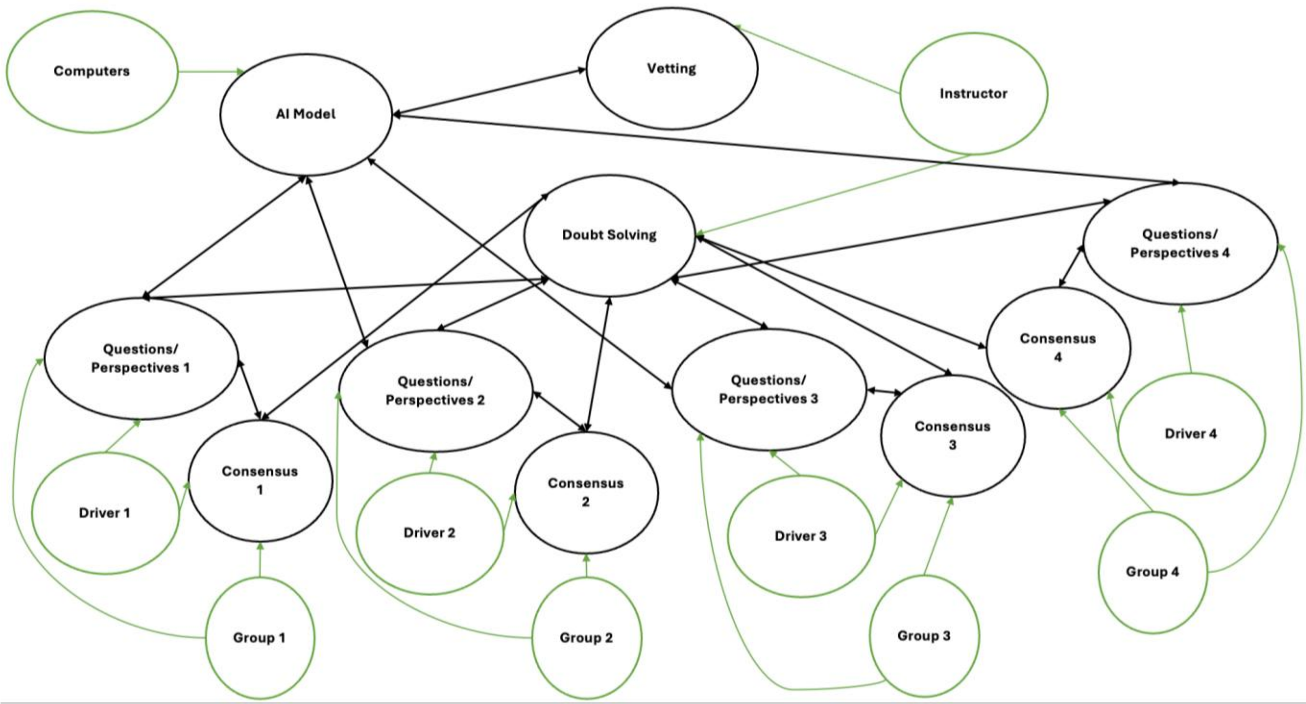


Note: *Green bubbles are M-individuals, and black bubbles are P-individuals in the diagram.*

The second blueprint (Figure 2) visualizes AI-mediated chat discussions. Before implementing these AI chats, the instructor interacted with the chatbot to ensure it worked smoothly and did not produce spurious information. During these sessions, two in number, conducted remotely over the course of half an hour towards the end of class (to enable efficient Breakout Room creation and prompting based on the preceding lecture), the instructor divided the class into four working groups. A representative "driver" prompted the AI chatbot after a short consensus exercise with their group of peers to decide upon the line of questioning to be followed.

After a brief conversation to set the stage, the student driver and peers asked the chatbot the weekly discussion prompt. The instructor hovered between the Breakout Rooms and solved students' doubts as needed.

Figure 2
Blueprint of AI-mediated chat.



Note: *Green bubbles are M-individuals, and black bubbles are P-individuals in diagram.*

5. The Current Study

This mixed methods action research study was co-authored by undergraduate students and their instructor and edited by two research assistants. Business-as-usual classroom discussions in a Google Chat community answering a weekly prompt are compared to AI-mediated group discussions held with chatbots representing eminent psychologists. Sociograms generated by RStudio's igraph package are used to compute the transitivity (incidence of three-way interactions between living and artificial systems), and average/participant eigen centrality (extent of egalitarian nature of agent participation and interconnectedness). Qualitative reflections detail group prompting mechanisms. The study answers two research questions:

RQ1: *To what extent do classroom discussions, when held in a conventional chat forum differ from those mediated by AI in the incidence of three-way interaction between living and artificial agents?*

RQ2: *To what extent do classroom discussions, when held in a conventional chat forum differ from those mediated by AI in the egalitarian participation of each living or artificial agent in conversational feedback loops?*

6. Method

Participant Observers

Nineteen college students and their instructor (20 in total, 50% Female, 50% Male, 35% White, 35% Black, 5% Pacific Islander, 10% Asian, 15% Mixed Race) acted as participant observers taking part in classroom discussions, collecting datapoints, analyzing them, and reflecting upon conducted activities. Two research assistants (50% Male, 100% Caucasian) working with the instructor only assisted in editing the paper after the study was completed. These two research assistants were part of an advanced internship class (PSY479) offered at the university, serving research contact hours for course credit. The study context was an introductory psychology class at a small liberal arts university in Southeastern Virginia. Each participant observer worked during six class sessions of 13 to implement the study as part of regular educational activities. The study was approved as an exempt project that reflected upon regular educational activities by the university's Institutional Review Board (IRB).

Curriculum

PSY101 was an introductory class for both majors in the field and those meeting general education requirements. The curriculum was drawn from Laura King's (2023) "*The Science of Psychology: An Appreciative View, 6th Edition*", published by McGraw Hill. The following topics were covered:

- **Chapter 1:** What is Psychology?
- **Chapter 2:** Research Methods
- **Chapter 3:** The Nervous System
- **Chapter 4:** Sensation and Perception
- **Chapter 5:** Human Consciousness*
- **Chapter 6:** Learning
- **Chapter 7:** Memory
- **Chapter 8:** Intelligence
- **Chapter 9:** Human Development*

The chapters were taught over two units. The first five were taught in Unit 1, and the last four in Unit 2. The class was graded based on several weekly assignments, two Unit assignments (to be prepared as a paper or storyboard), and a group final. At the beginning of each class, Google chat discussions were held about the weekly topic. Two chats were held using AI chatbots (sessions are marked with an asterisk in the bulleted list above) on the Character.AI and DeepAI platforms. Business-as-usual chats were followed by the lecture covering the topic, and students would complete a McGraw Hill Connect Smartbook worksheet at home as asynchronous work each week. AI chats were conducted after the session lectures (held remotely) to enable student groups and their representative “driver” to question the chatbot effectively in Breakout Rooms.

The class worked on a project-based task involving comparative analyses and reflections on AI and non-AI discussion boards, which culminated in the present manuscript. Two workshop style sessions were held, during which the paper was edited by the whole class, and final presentations were worked on.

Data

Data were drawn from four of the discussion boards held during class, spanning between 20-30 minutes. Business-as-usual chats were conducted on a Google community with live threaded posting functionality. The instructor collected business-as-usual chat data by initially pasting it into a cloud-based document. Chats on the nervous system and human learning were considered for analysis, as they occurred in proximity to the AI chats. The experimental AI conversations were held on the Character.AI platform and DeepAI platforms. Students and the instructor discussed the weekly prompt in online working groups with a chatbot representing Freud and Piaget, during sessions on human consciousness and development. Students copy-pasted their AI discussion questions and responses into a Google document for further processing.

Conversations between the group members were not recorded by external observers to maintain intramundane participant observation. Each group answered a rubric asking about the process followed as a group to reach consensus in prompting AI-mediated conversations, generating qualitative data used to describe these conversations.

Measures

RStudio Igraph Package: The instructor collected data from the four chats and inputted interactions occurring in them between students, the instructor, and language models part of the AI discussions into a Microsoft Excel sheet as edge lists. These four edge lists were analyzed in RStudio using the igraph package. Network analysis allows an understanding of the mechanisms guiding community formation and distributed talk and has been widely used in contemporary social science research (Kolaczyk & Csárdi, 2014).

Reflection Rubric: During the two AI-mediated chat sessions, students were divided into four working groups that prompted chatbots of Freud and Piaget after brainstorming a cogent line of questioning. However, in our network analysis, only the student driving the prompting is considered, since group conversations in Breakout Rooms were not video recorded. After completing the conversation, the whole group reflected on the seven question rubric provided below to explain the consensus building process and the conversation the group and chatbot had. This qualitative data source accounts for possible lapses in our quantitative analyses. While word counts were suggested for each answer in the rubric, the instructor allowed them to be open-ended and asked students to answer the questions to their best capacity:

1. How did you prompt AI initially to converse with [bot name] as a group? Describe how you conversed to come up with your initial questions to [bot name] in 100 words or more.
2. Did the AI always respond correctly? What did you do to redirect it if it did not? Answer in 50 words or more.
3. When you asked [bot name] the discussion prompt, how did it respond? How did you further converse with it after? Describe in 50 words or more.
4. What conclusion did you reach with [bot name] at the end of the chat? Describe in 50 words or more.
5. How did the bot argue with you, if at all? Describe how it influenced the conversation in a sentence or two.
6. Think about your Google Chats that we do weekly in class as a group. How is this different? Answer in 100 words or more.
7. How does adding AI to a discussion activity change it? Answer in two sentences or more.

Data Analysis

A mixed methods approach is used. Edge lists of all four discussions exported from Microsoft Excel as a .csv file were analyzed in RStudio using the igraph package. Network sociograms were plotted for recorded interactions in each discussion along with metrics of transitivity and eigen centrality in each conversation (Kolaczyk & Csàrdi, 2014). Transitivity measures the proportion of interactions in the class network occurring at a three-agent level or more. Eigen centrality helps understand the incidence of egalitarian connections between agents having different levels of activity in the network.

While transitivity and eigen centrality were computed on average for the whole network, the reliance on individual students versus drivers in each network analysis dramatically reduced the number of ties in the AI data. This prompted the additional computation of individual eigen centrality of student drivers in the non-AI and AI chats to allow for a more cogent comparison. Word counts of questions and answers, the average degree (communications to and from each agent) and the number of conversational turns were also computed.

To overcome limitations in data collection, our comparative analysis of network sociograms is supplemented by a narrative inquiry (Connelly & Clandinin, 1990) of students' reflection rubric responses that compared the mechanisms at play during AI and non-AI chats and described group prompting mechanisms.

7. Results

In our sociograms of all chats, pink nodes represent agents engaging in conversational turns greater than or equal to the average degree of the network, while blue nodes are agents with below average degree. Node size is directly proportional to degree. Edge width is proportional to the weight of the turns between agents part of an edge. Tying the network key back to the M- and P-individual framework, each node represents comments posted by an agent (one or more P-individuals embodied in an M-individual), and thus becomes an abridged version of an M- and P- blueprint.

Both the business-as-usual discussions were held on the class's Google Chat community. The topics considered for the business as-usual-chats used as comparison for our AI chats focused on the human nervous system, and education/learning. These chats were proximal in time, occurring only one to three weeks prior to each of the considered AI chats, and focused on topics comparable to the AI discussions.

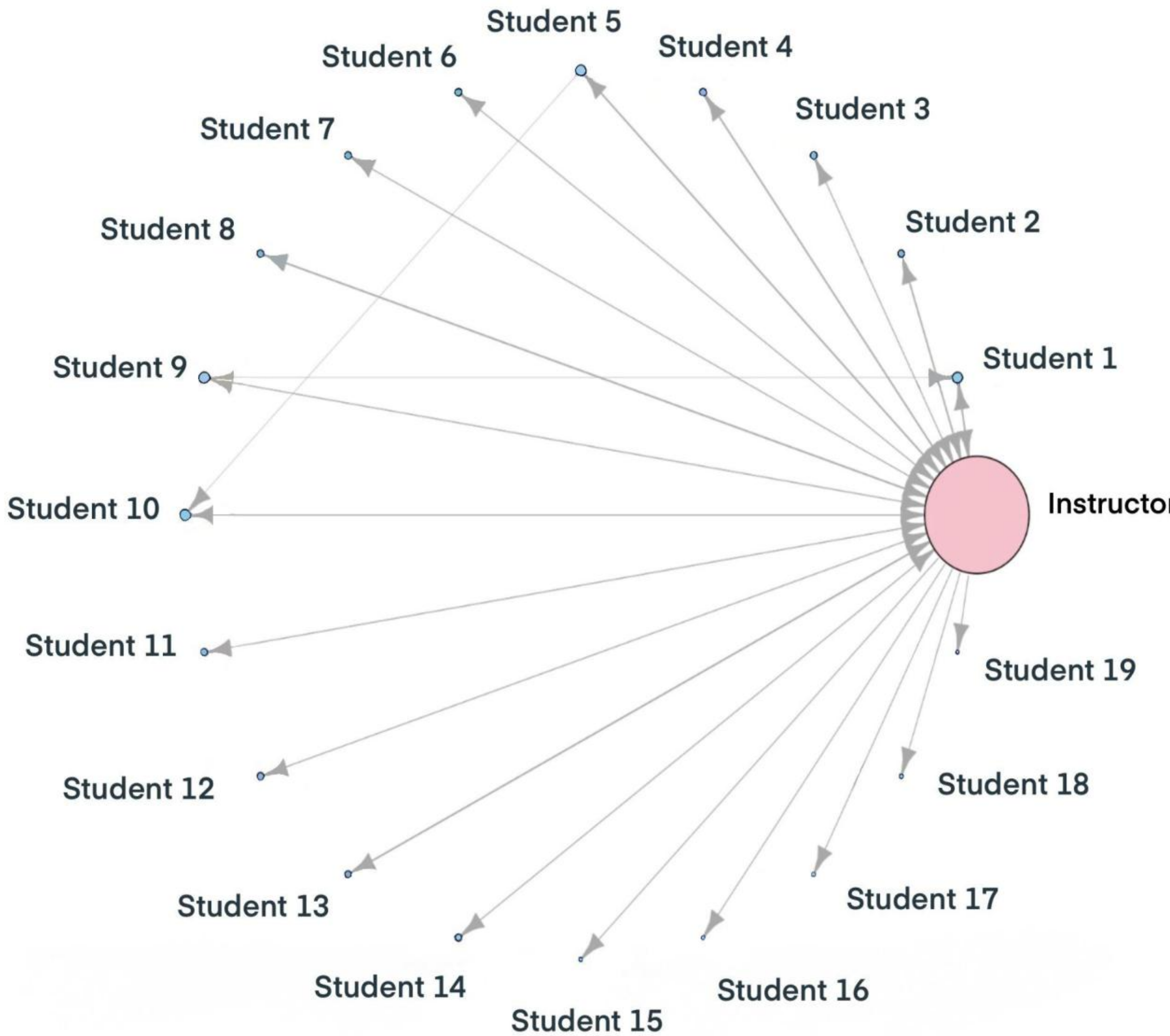
Network sociograms for both business-as-usual chats showed that the interactions ensuing during these chats were dominated by student responses to the instructor's weekly prompt.

Only a sparing amount of student replies to their peers were observed in the chat focused on the nervous system during Lecture 3, in which a total of 14 students responded to the prompt (Figure 3). The prompt asked:

What is the origin point of learning differences like autism spectrum disorder and ADHD? Is it owing to a difference in the structure of the nervous system, social experience or both? State your position and your "why" and "how".

Figure 3

Google Chat discussion covering the nervous system.



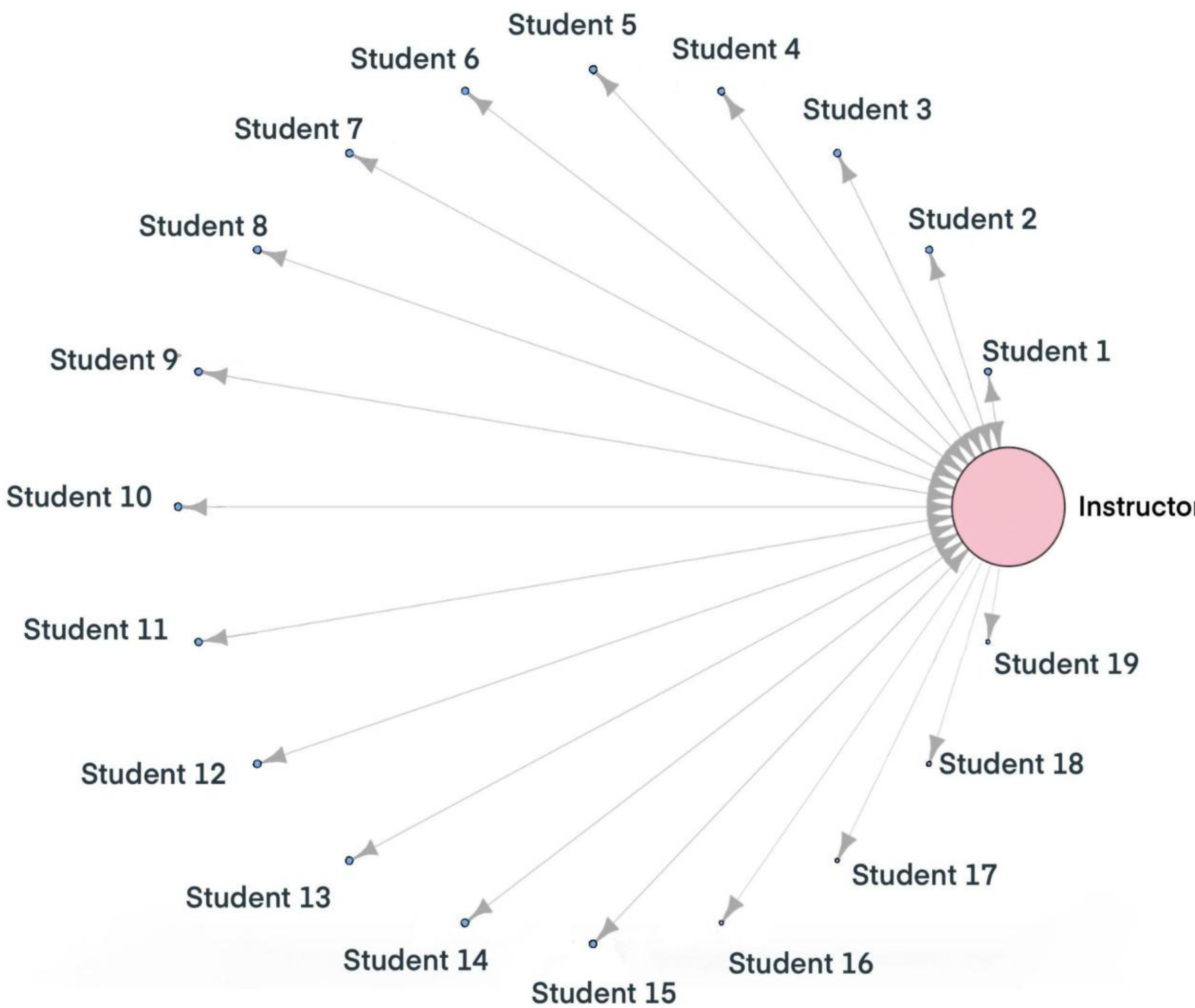
In the second business-as-usual chat (Figure 4) on learning, during Lecture 6, the prompt was as follows:

How did your teachers use reinforcement and punishment in the classroom in your K-12 experience (up to high school)? How did it affect your learning?

While posts from each student were rich and recounted their educational histories, there were no efforts made by individual learners to comment on their peers' perspectives, showcasing how a highly individualistic mode of response was seen.

Figure 4

Google Chat discussion covering learning.



Compared to the business-as-usual chats, the AI-mediated chats on human consciousness and development were conducted differently. Groups of students having to interact with chatbots of Freud and Piaget. Classes were held online. The four working groups collectively brainstormed how to prompt the chatbot for the week and begin a conversation about the weekly topic, engaging in four concurrent conversations.

Each group was assigned a Breakout Room on Google Meet. Members chose a “driver” to prompt the bot and share their computer screen. This student would input the questions that the group would reach consensus on.

In the AI-assisted chat on human states of consciousness in Lecture 5, 16 students engaged in a conversation with a chatbot representing Freud on the Character.AI platform, and spoke to it about the varied reasons and explanations for dreaming, followed by a critique of Freud’s perspectives on the roots of dream phenomena by asking about human behavior and its emergent nature. Before starting, the instructor made sure the platform was working fine by interacting with the chatbot and verified that it would not produce false information.

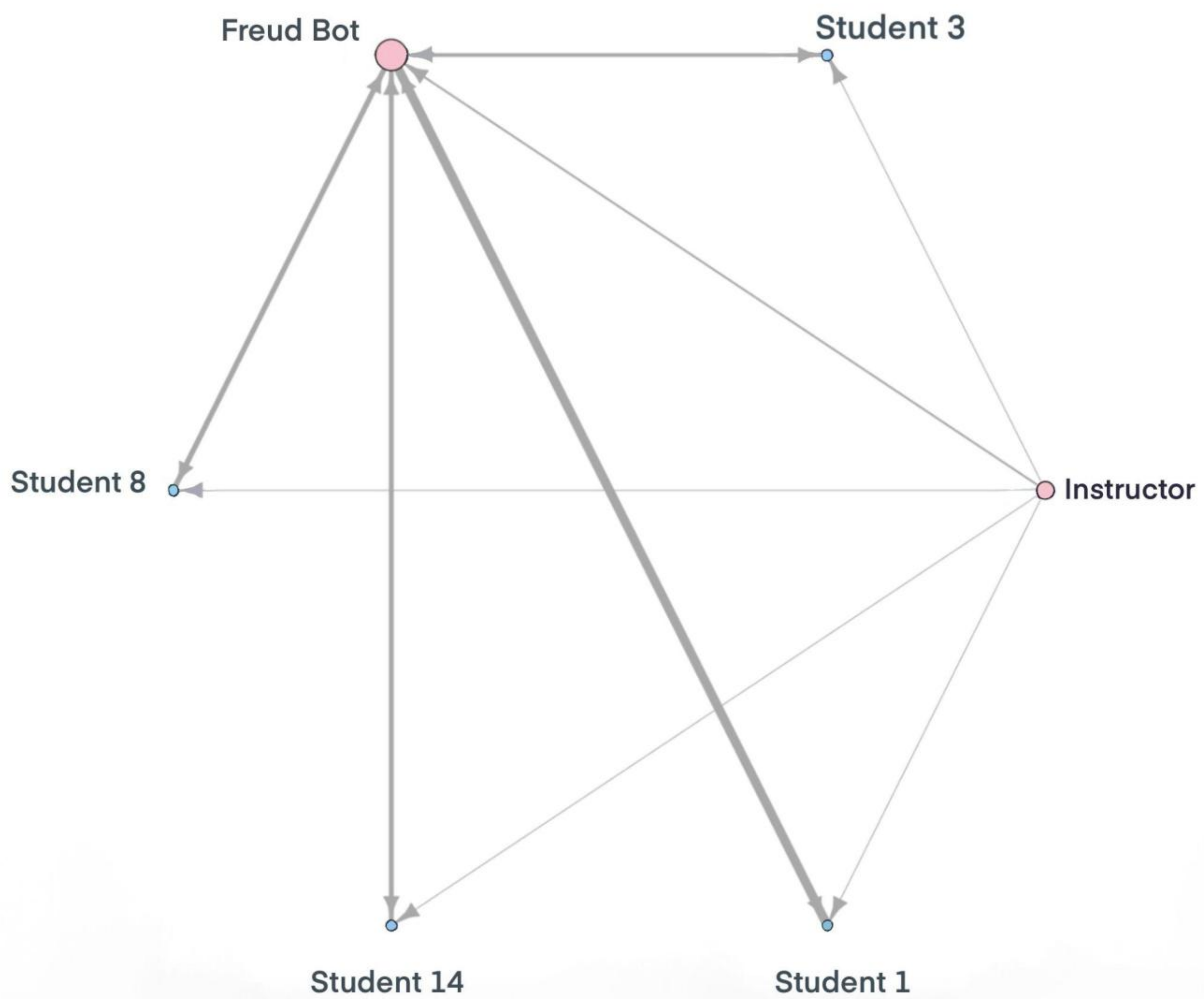
Students brainstormed prompts together, and the driver input them into the platform. The agents engaged in repeated conversational turn taking with the chatbot, and were also guided by the instructor, who hovered from group to group to shepherd the discussions twice. After conversing with the Freud bot, the student drivers were requested by the group to ask it the weekly prompt:

How does the concept of the unconscious mind explain seemingly irrational behaviors in adults? What are the limitations of using this approach?

Students 1, 3, 8 and 14 played the role of drivers in the first AI chat. The nodes representing the group leaders are presented in the network diagram below (Figure 5), but the conversations between other members of each group and the drivers are captured only through qualitative reflections. Students and instructor decided not to record class sessions, or invite observers to maintain a purely intramundane source of observation. This decision was made to avoid the small chance for the Hawthorne Effect to influence classroom activity, as we have mentioned in the Theoretical Framework of this paper.

Figure 5

AI-mediated chat on human consciousness.



During the class on human development in Lecture 9, a second AI-mediated chat was conducted (Figure 6). Fourteen students participated. They worked in four groups and interacted with a language model of Piaget on the DeepAI platform, with the instructor hovering between groups to assist three to five times.

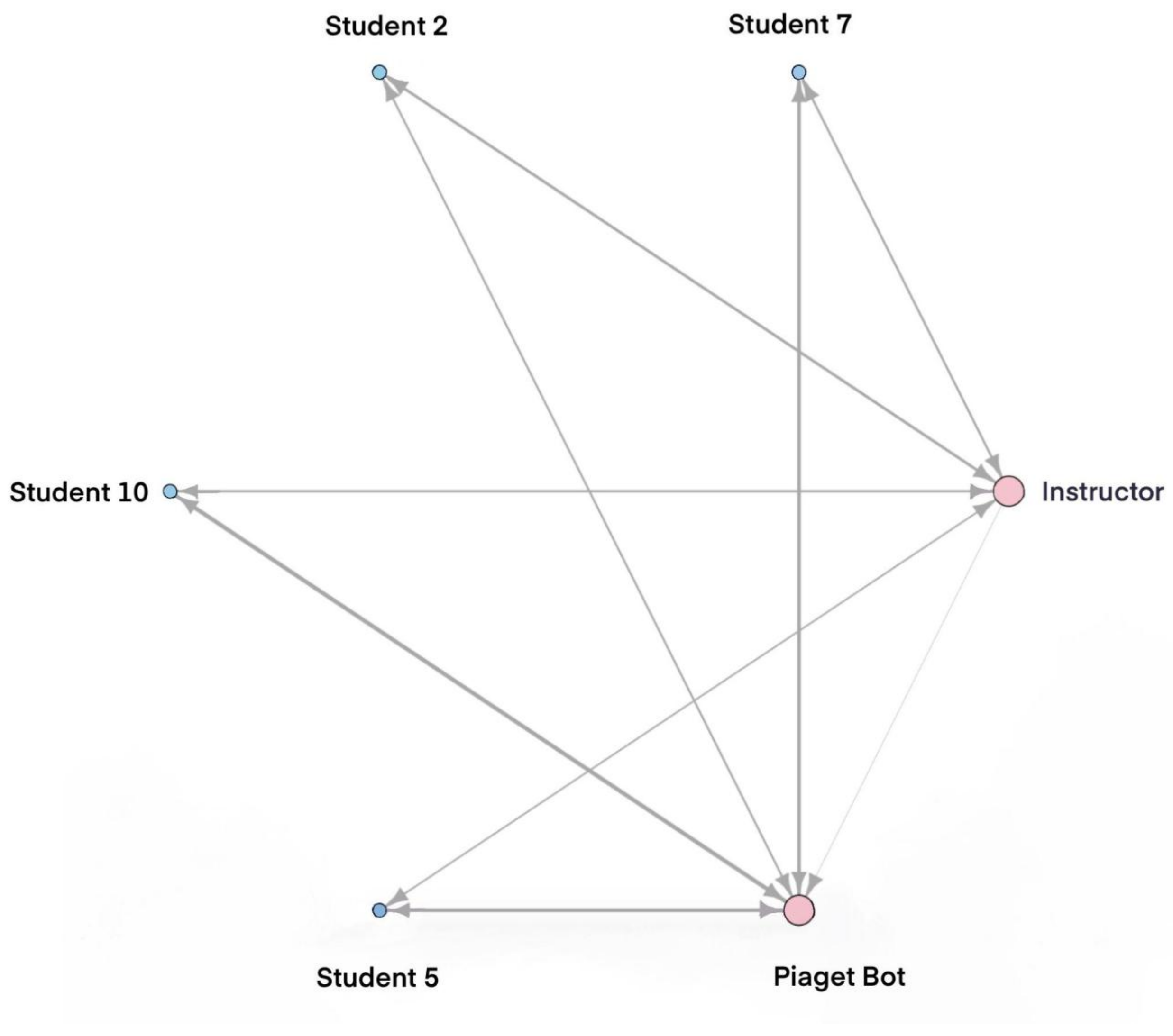
A similar process was followed in each group to shepherd a conversation with the Piaget bot, with one student in each group driving the prompting. Students 2, 5, 7 and 10 acted as the drivers.

After briefly interacting with the chatbot, the group and the representative driver asked bot the weekly prompt:

How do assimilation and accommodation play a role in cognitive development? Show how the process occurs using an everyday example.

Figure 6

AI-mediated chat on human development.



Metrics such as transitivity, eigen centrality, the number of conversational turns, total words (skewed by the long responses of the two bots), and average degree or inward and outbound interaction for conversational agents in each AI chat were seen to be higher than in the business-as-usual discussions. The fully interpreted metrics from each chat are provided below in Table 1. These metrics serve to answer both RQ1, and 2, which asked about egalitarian participation and three-way talk in AI and non-AI chats.

Table 1

Comparative analysis of AI and non-AI chats.

Topic	Technology	Turns	Words	Avg. Eigenvector Centrality	Transitivity	Avg. Degree
Nervous System	Google Chat	60	1420	0.246	0.035	3.5
States of Consciousness	Character AI	94	2580	0.63	0.5	4.33
Learning	Google Chat	34	1159	0.263	0	3.4
Development	Deep AI	81	10338	0.764	0.5	5.67

The higher incidence of transitive interactions is only partially depicted in our network diagrams, since the activity of each group subsumes collective brainstorming between multiple students, input by the driver to the bot. Our results may not accurately depict multi-agent conversations within each decentralized group, but provide a rough picture of the observable collected human-AI interactions. Concerns with our analysis arise from the consideration of the driver as prompter. This reduces the total number of recorded feedback loops, which may increase transitivity (since transitivity is essentially a probability or percentage score). Capturing descriptions of these conversations using alternative methods, and even metrics for individual students can ensure that graphical results are not heavily skewed in a positive direction through a reduction of the total number of edges in the AI networks.

In Table 2, the eigen centrality of the individual students acting as drivers in the AI chats is compared across our business as usual and experimental configurations to provide a more equivalent comparison between the two types of chat networks. Individual eigen centrality metrics contribute towards better answering RQ2.

Table 2

Comparison of individual eigen centrality in business as usual and experimental chats.

Student ID	Non-AI eigen centrality	AI-mediated eigen centrality
Student 1	0.29	0.54
Student 2	0.25	0.65
Student 3	0.25	0.54
Student 5	0.29	0.65
Student 7	0.25	0.65
Student 8	0.25	0.54
Student 10	0.29	0.65
Student 14	0.25	0.54

Since the number of agents in each setting is slightly different (group drivers vs. each student) and may not capture all conversational feedback loops in the AI chats, we also recount mechanisms at play among the human agents in each of the four groups to support quantitative results. Qualitative reflections helped gauge if increased three-way interactions owing to treatment of the student driver as a node representing the group did not skew our analysis, and if there was truly a decentralized group consensus process preceding prompting.

We narrate the insights of each group engaging in AI-mediated discussions that express how students perceive these chats as different from conventional forums and describe their group consensus procedure. Group 4's students, in their response, stated:

"The interaction with the AI is more descriptive and precise with the questions that the AI receives, rather than interacting with other students giving logical answers and knowledge. Some information may not be more accurate than the AI but it's more authentic."

This appreciation of cold hard facts provided by the AI model, so that students could learn about psychological theories and grapple with complex ideas rather than arguing with peers right at the outset, was also expressed by Group 3 in a more positive light:

"We were receiving information directly from the "real" scientist rather than relying on prior knowledge. He consistently provided thorough answers, supported by facts, making it feel more credible since it appeared to come directly from him. It changes it by making the content feel more authentic, as if you're hearing directly from a real scientist. This made it more engaging because it was easy to follow and didn't feel argumentative."

While some groups of students expressed greater comfort with the facts provided by the bot, and its adherence to a standpoint matching its theory, others, such as those in Group 1 wished for the bots to have capacity to falsify their own standpoint, saying:

"He [Freud] was very repetitive and persistent about his own opinions and ideologies. We continued to ask further questions that questioned his theory and some bias that surrounds it. His responses still revolved around his own theory and did not sway away from what he thought was correct."

The process followed by each group contributed to the interactive, dynamic nature of the AI-mediated chats. The use of the collective "we" in all reflection responses is the first indicator of collaborative group talk guiding the driver's prompting. Student co-agency led to an organic building up of questions upon each other. The process often began spontaneously, highlighting that the conversations guiding the driver's prompting were emergent and goal-oriented. Per Group 2's experiences chatting with the Piaget bot:

"To start talking with Piaget, we just picked a topic we were interested in like," What fosters attachment?" and we asked him to give us some examples. Then, we asked other questions to see how he might answer, based on what we know about his ideas. After that, we asked more questions to keep the conversation going and understand his theory better. It was a simple way to learn by pretending to talk to him and seeing how he would explain things in his own style."

Group 4's students also shared similar insights about their chat with the Freud chatbot, recounting how they relied on concepts they learnt in the lecture to draft questions as a group, solidifying the organic progression between the lecture and the chat:

"Our group initially conversed with Freud by asking his own thoughts and opinions to get a baseline for what is to be considered his theory. We conversed to come up with our initial questions by taking points from our slideshow and previous lecture and turning them into specific questions that we were more curious to learn deeper about."

Narrated insights from the reflection rubric showcase the group talk between the student drivers and their peers that were not recorded for inclusion in our network analysis of AI-mediated chats.

Our mixed methods analysis compensates for a lack of detailed data about group prompting that we interpreted using network statistics derived from only the student driver's activity in the AI-mediated chats. The role of group consensus in distributed AI-mediated interaction, and the involvement of each agent in the decision-making process expressed in student reflections support the idea that AI-mediated chats involved a greater degree of decentralized collaboration.

8. Discussion

This mixed methods action research study focuses on the use of GenAI as a Socratic Opponent in classroom discussions. It answers two research questions. RQ1 asks: *To what extent do classroom discussions, when held in a conventional chat forum, differ from those mediated by AI in the incidence of three-way interaction between living and artificial agents?* Unlike the minimal transitive interactions in our business-as-usual Google community chats, the AI-mediated group discussions showed greater transitivity. This was largely due to the presence of a surrogate conversational agent (the AI) that prompted continued questioning and deeper engagement, as well as the instructor's active role in facilitation and doubt-solving. The consensus formation process followed by each set of students shared through written qualitative reflections also highlights the co-agency of human agents in the system and ensures that the sole consideration of the driver as prompter does not skew results too much. The higher number of conversational turns and averaged degree of each node also highlight how interactions and social engagement were richer in the AI chats.

RQ2 asks: *To what extent do classroom discussions, when held in a conventional chat forum differ from those mediated by AI in the egalitarian participation of each living or artificial agent in conversational feedback loops?* The higher average eigen centrality of nodes in the AI chat networks indicates the shared participation of agents (the groups, AI bots, and the instructor) in the conversation. The higher eigen centrality of individual student drivers in the AI-mediated networks further support our results, and accounts for the fact that the nature of recorded data used to generate both networks is slightly different. The aspects of group level feedback loops missed out in our AI-mediated chat analysis were compensated for by the reflections of students, which highlighted the egalitarian talk in developing prompts. The role of the instructor in hovering from group to group to assist in problem-solving also contributed to the interactive nature of the conversation.

The present study takes on a different approach in comparison to extant efforts where AI agents are the guides of Socratic reasoning (Blasco & Charisi, 2024), often producing tangentially directed conversations, albeit with high social engagement. Students questioned chatbots in elaborating upon theories and concepts, vetting information and ensuring prompts build on one another productively as a group. The presence of the instructor two to five times in each group to help facilitate conversation (as opposed to an AI prompting increased participation; Do et al., 2022), data collection and reflection adds a participatory flavor to the study, the product of which is the current paper. Thirdly, the lecture preceding the AI-mediated chat lowered the likelihood that students would let the AI provide them with false information that would detract from their knowledge. Group 4's qualitative reflections highlighting the role of the lecture and slide content in helping ensure accuracy of content reinforced this notion. Furthermore, upon checking the contents of the chats, no spurious information was found.

Our network analysis adds a new dimension to existing research that has so far, focused on calculating the frequency of like-based and commenting behaviors in AI-mediated settings (Haqbeen et al., 2023). Overall, our study forms a novel contribution that helps understand how to use AI to add a dynamic component to a college classroom discussion and compares the nature of conversational networks in traditional and AI-mediated chats.

9. Limitations

There are a few limitations of this study. The first is the small sample size and the consideration of a single classroom setting. However, a mixed methods analysis that relies on qualitatively expressed experiences, and network metrics overcomes this limitation. We provide rich data from a small classroom community optimal for creating modular, decentralized conversations (Smith et al., 2020).

The second limitation is the nature of the network analysis of the AI chats. The only students considered in each group in the network analysis were single student drivers inputting prompts to the AI chatbots after forming consensus. However, the brainstorming process itself, between multiple group members was not analyzed using network metrics. These logistical issues became inevitable owing to the discontinuation of the group chat feature in Character.AI, which earlier allowed multiple human agents to talk to one chatbot. The change in available features led to our choice to experiment with both Character.AI and DeepAI and have one single student prompt the bots after a group consensus exercise.

The disparity in our network analysis leads to the production of visualizations depicting a highly concentrated set of conversational feedback loops in the AI-mediated chat sociograms.

Quantitatively capturing the conversations at play during the group brainstorming process in the online AI chats could have only been accomplished by recording each Breakout Room or inviting observers. We forewent this opportunity for an ethnomethodological analysis (Haataja et al., 2022) owing to the possibility for awareness of observation to change student activity, ir a Hawthorne Effect (Oswald et al., 2014). However, if an analysis of video recordings/observed talk had been implemented, the average eigen centrality of each node in the AI chat networks would be different. We attempt to overcome this disparity through a comparison of eigen centrality for each student driver across the two types of chat configurations. Moreover, qualitative explanations of group processes help highlight how consensus and distributed talk produced a cogent line of questioning with each AI chatbot.

A third limitation is the implementation of the AI-mediated discussion at the end of the classroom lecture. This setup led to students appreciating the cold, hard facts provided by AI, but also saying that both classroom argumentation and machine-mediated talk are valuable and can reinforce each other. Future suggestions to design the activity involve leaving some time at the end of class to facilitate a classroom discussion between instructor and peers to further reflect on AI-mediated talk. The activity will be restructured in this manner for future semesters, consonant with the principles of participatory action research (Glassman et al., 2012), and Pask's cybernetics (De Zeeuw, 2001).

The scrappy nature of this action research study is emblematic of the nimbleness in cybernetic research, even though it may sacrifice the extramundane source of truth that conventional scientists rely on, and veer from the language of generalizability. Instead, our study takes the example of a highly targeted college classroom and showcases how both learners and instructors were able to reflect upon the activities they engaged in, rather than completing assignments for assignment's sake.

10. Conclusion

With generative AI becoming increasingly embedded in work, educational and informal activity trajectories for humans in the Information Age, opportunities to explore its uses and stretch possibilities to their limit abound. This study forms a targeted response to Mike Sharples' (2023) call for practical implementation of contemporary AI tools in distributed educational ecologies. It specifically focuses on assigning the role of the Socratic Opponent to Character.AI and Deep AI chatbots. Takeaways indicate that AI can provide cold hard facts within a constrained theoretical standpoint to help students master and grasp psychological concepts, and defend theories when critiqued. However, both machine dependent and independent classroom activities each have their place in facilitating learning, highlighting the value in an undulation between human led and AI-mediated problem-solving. By adopting the principles of second-order cybernetics, our study establishes a bilingual sensibility of technology usership (Pangaro, 2021). Per this sensibility, human agency and technology symbiotically feed one another; producing emergent insights to enable cogent machine-mediated problem-solving.

References

Behymer, K. J., & Flach, J. M. (2016). From autonomous systems to sociotechnical systems: Designing effective collaborations. *She Ji: The Journal of Design, Economics, and Innovation*, *2*(2), 105-114. https://doi.org/10.1016/j.sheji.2016.09.001Get rights and content

Blasco, A., & Charisi, V. (2024), The Impact of Large Language Models on Students: A Randomised Study of Socratic vs. Non-Socratic AI and the Role of Step-by-Step Reasoning. Social Science Research Network (SSRN). http://dx.doi.org/10.2139/ssrn.5040921

Bostrom, N. (2014). *Superintelligence: paths, dangers*. Oxford University Press.

Connelly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational researcher*, *19*(5), 2-14. https://doi.org/10.3102/0013189X019005002

De Zeeuw, G. (2001). Interaction of actors theory. *Kybernetes*, *30*(7/8), 971-983. https://doi.org/10.1108/03684920110396864

Do, H. J., Kong, H. K., Lee, J., & Bailey, B. P. (2022). How should the agent communicate to the group? Communication strategies of a conversational agent in group chat discussions. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1-23. https://doi.org/10.1145/3555112

Fakour, H., & Imani, M. (2025). Socratic wisdom in the age of AI: a comparative study of ChatGPT and human tutors in enhancing critical thinking skills. *Frontiers in Education*, *10*, 1528603. https://doi.org/10.3389/feduc.2025.1528603

Fleer, M. (2016). Theorising Digital Play: A Cultural-Historical Conceptualisation of Children's Engagement in Imaginary Digital Situations. *International Research in Early Childhood Education*, *7*(2), 75-90. https://files.eric.ed.gov/fulltext/EJ1138705.pdf

Glassman, M., Bartholomew, M., & Hur, E. H. (2013). The importance of the second loop in educational technology: An action science study of introducing blogging in a course curriculum. *Action research*, *11*(4), 337-353. https://doi.org/10.1177/1476750313502555

Haataja, E. S. H., Chan, M. C. E., Salonen, V., & Clarke, D. J. (2022). Can noncomplementarity of agency lead to successful problem solving? A case study on students' interpersonal behaviors in mathematical problem-solving collaboration. *Learning and Instruction, 82*, Art. 101657. https://doi.org/10.1016/j.learninstruc.2022.101657

Halpin, H. (2025). Artificial intelligence versus collective intelligence. *AI & Society.* https://doi.org/10.1007/s00146-025-02240-x

Haqbeen, J., Sahab, S., & Ito, T. (2022, June). How did discourse shift among Afghan citizens during the fall of republic: Early insights using conversational AI platform. In *Proceedings of the 23rd Annual International Conference on Digital Government Research* (pp. 480-482). https://doi.org/10.1145/3543434.3543595

Held, P., Pridgen, S. A., Chen, Y., Akhtar, Z., Amin, D., & Pohorence, S. (2024). Mixed method feasibility study protocol for Socrates 2.0: a novel cognitive behavioral therapy-based generative ai tool to facilitate Socratic dialogue. JMIR Research Protocols, 13. 10.2196/58195

Kim, S., Eun, J., Oh, C., Suh, B., & Lee, J. (2020, April). Bot in the bunch: Facilitating group chat discussion by improving efficiency and participation with a chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13). https://doi.org/10.1145/3313831.3376785

King, L. A. (2023). *The science of psychology: An appreciative view*. McGraw Hill LLC.

Kolaczyk, E. D., & Csárdi, G. (2014). *Statistical analysis of network data with R* (Vol. 65). Springer.

Licklider, J. C. (1960). Man-computer symbiosis. *IRE transactions on human factors in electronics*, (1), 4-11. 10.1109/THFE2.1960.4503259

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4). https://doi.org/10.1609/aimag.v27i4.1904

Oswald, D., Sherratt, F., & Smith, S. (2014). Handling the Hawthorne effect: The challenges surrounding a participant observer. *Review of social studies*, *1*(1), 53-73. https://www.rossjournal.co.uk/wp-content/uploads/2016/12/RoSS-Vol1-No1-Oswald-et-al-53-73.pdf

Pangaro, P. (2021). Responding to the Pandemic of Today's AI. Design + Conversation. Retrieved From: https://pangaro.com/designconversation/2021/08/newmacy-in-2021-pandemics-ai/

Pask, G. (1966). Comments on the cybernetics of ethical, psychological and sociological systems. In J. P. Schade (Ed.), *Progress in biocybernetics* (Vol. 3, pp. 158–250). Elsevier. https://www.pangaro.com/pask/pask%20comments-cyber-ethical-social-psychological-systems.pdf

Pask, G. (1975). *Conversation, cognition and learning: cybernetic theory and methodology*. Elsevier.

Pask, G. (1976). *Conversation theory, applications in education and epistemology*. Elsevier.

Reinertsen, A. B. (2012). Second order pedagogy as an example of second order cybernetics. *Reconceptualizing Educational Research Methodology*, *3*(1). ttps://doi.org/10.7577/rerm.358

Rubtsova, O. V., & Salomatova, O. V. (2022). Child's play in the context of digital transformation: cultural-historical perspective (part one). *Cultural Historical Psychology*, *18*(3), 22-31. https://doi.org/10.17759/chp.2022180303

Sabzalieva, E., & Valentini, A. (2023). *ChatGPT and artificial intelligence in higher education: Quick start guide*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000385146

Scott, B. (2016). Reflections on the sociocybernetics of social networks. Social Science Research Network (SSRN). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2883537

Sharples, M. (2023). Towards social generative AI for education: theory, practices and ethics. *Learning: Research and Practice*, *9*(2), 159-167. https://doi.org/10.1080/23735082.2023.2261131

Smith, N. R., Zivich, P. N., Frerichs, L. M., Moody, J., & Aiello, A. E. (2020). A guide for choosing community detection algorithms in social network studies: The question alignment approach. *American journal of preventive medicine*, *59*(4), 597-605. 10.1016/j.amepre.2020.04.015

Tilak, S., Bagley, B., Cantu, J., Cosby, M., Engelbert, G., Hammonds, J. K., ... & Williams, M. (2024b). Using text-to-image generative AI to create storyboards: Insights from a college psychology classroom. *Journal of Sociocybernetics, 19*(1), 1-42. https://doi.org/10.26754/ojs_jos/jos.2024110822

Tilak, S., & Bogacki, M. (2024, June). Creating a cybernetic feedback mechanism using Assessment and Learning in Knowledge Spaces (ALEKS) for high schoolers with ADHD in mathematics classrooms. *Presented at the American Society for Cybernetics 60th Anniversary Meeting, Washington, DC.*

Tilak, S., & Glassman, M. (2022). Gordon Pask's second-order cybernetics and Lev Vygotsky's cultural historical theory: Understanding the role of the Internet in developing human thinking. *Theory & Psychology*, *32*(6), 888-914. https://doi.org/10.1177/09593543221123281

Tilak, S., Glassman, M., Kuznetcova, I., & Pelfrey, G. L. (2022). Applications of cybernetics to psychological theory: Historical and conceptual explorations. *Theory & Psychology*, *32*(2), 298-325. https://doi.org/10.1177/09593543211053804

Tilak, S., Lincoln, J., Miner, T., Christensen, N., Jankowski, J., & Kennedy, K. (2024a). A participatory, qualitative analysis of the use of MagicSchool AI for course design. *Journal of Sociocybernetics*, *19*(1), 43-106 https://doi.org/10.26754/ojs_jos/jos.2024110947.

Tilak, S., Manning, T., Glassman, M., Pangaro, P., & Scott, B. (2024c). Gordon Pask's conversation theory and interaction of actors theory: Research to practice. *Enacting Cybernetics*, *2*(1). https://doi.org/10.58695/ec.11

Tzafestas, S. G. (2017). *Systems, cybernetics, control, and automation*. River Publishers.

Westermann, C. (2018). On delight: Thoughts for tomorrow. *Technoetic Arts: A Journal of Speculative Research*, *16*(1), 43-51. https://doi.org/10.1386/tear.16.1.43_1

