Dg.o2025 - 26th Annual International Conference on Digital Government Research
09 – 12 June, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre City, Brazil

Assessment of the LLM-based Chatbots on Student Engagement and Learning Outcomes in Afghanistan.

Jawad Ahmad Haqbeen*, *Sofia* Sahab, *Takayuki* Ito

Graduate School of Informatics, Kyoto University, Kyoto, Japan, email address: jawad.haqbeen@i.kyoto-u.ac.jp, ORCID 0000-0002-0481-0196

Graduate School of Informatics, Kyoto University, Kyoto, Japan, email address: sahab.sofia@i.kyoto-u.ac.jp, ORCID 0000-0002-5822-5021

Graduate School of Informatics, Kyoto University, Kyoto, Japan, email address: ito@i.kyoto-u.ac.jp, ORCID 0000-0001-5093-3886

Submitted: 31 January 2025, Revised: 26 March 2025, Accepted: 21 April 2025, Published: 19 May 2025

Abstract. The integration of Generative AI (GenAI) technologies, such as ChatGPT, into online education is accelerating; however, their effectiveness in under-resourced contexts remains insufficiently studied. This paper investigates the impact of a Large Language Model (LLM)-based conversational agent on student engagement and learning outcomes in Afghanistan, where access to formal education—particularly for women—is severely restricted or banned. We conducted an experimental study involving 80 undergraduate computer science students (40 male, 40 female) in Afghanistan, randomly assigned to control and treatment groups. All participants attended identical 50-minute online lectures followed by 40-minute post-lecture discussions moderated by a human instructor, and completed a follow-up self-report questionnaire. The treatment group additionally engaged in AI-facilitated discussions using a GPT-4-based chatbot during post-lecture discussion. Analysis of discussion logs and post-intervention surveys revealed that the treatment group demonstrated significantly higher participation rates, with more posts and replies, during post-lecture discussion and reported greater confidence in their understanding of the course material. These findings highlight the potential of LLM-based chatbots to enhance interactive learning and foster educational inclusion, particularly for marginalized populations in low-resource environments.

Keywords. Conversational AI, online deliberation, online learning, GPT-4, LLMs, AI, ethnographic studies, learning outcomes, women, Afghanistan
Research paper, DOI: https://doi.org/10.59490/dgo.2025.956

1. Introduction

The demand for online learning has surged globally due to limited classroom capacity, rising educational costs, and the need for flexible learning options (Hart et al., 2018) External disruptions such as the COVID-19 pandemic (Szopiński & Bachnik, 2022) and internal political and cultural restrictions in such war-ravaged countries as Afghanistan have further accelerated the shift toward online education (Haqbeen et al., 2023). Afghanistan is currently the only country where women and girls are systematically banned from accessing secondary and higher education due to political and cultural restrictions imposed by the ruling Taliban regime (Haqbeen et al., 2023). Recent surveys in Afghanistan highlight the severe impact of these restrictions on women's educational attainment, with females significantly lagging behind males in learning and academic achievement (Lyons, 2023). This disparity reflects the longstanding gender inequality in access to education in Afghanistan (Shayan, 2015), where women and religious minorities are frequently excluded from male-dominated spheres such as education, politics, commerce, and even public life (Haqbeen et al., 2021).

In response, many Afghan women now rely on online learning opportunities provided by non-profit organizations as their primary avenue for education (Haqbeen et al., 2023). Despite the promise of online education, virtual learning environments often suffer from limited interactivity and student engagement (Singh & Meena, 2022). A lack of meaningful student–student, student–instructor, and student–content interaction during online course is a major drawback, leading to passive learning experiences (Haleem et al., 2022). Large class sizes (sometimes with hundreds of participants) pose additional challenges for maintaining engagement and personalized interaction. Studies have found that increasing online class size can negatively impact instructor–student interaction quality (Sorensen, 2015), and reduce student satisfaction (Russell & Curtis, 2013). Indeed, research shows that interactivity—especially between students and instructors—is a key factor influencing student satisfaction and learning outcomes (Espasa & Meneses, 2010). Active teaching strategies that promote engagement have been shown to improve participation (Deslauriers et al., 2019). Facilitating peer-to-peer learning is considered crucial for the next wave of online education innovation (Chandra & Palvia, 2021). However, in large online classes it is difficult for instructors alone to manage and simulate rich post-lecture discussions among students (student–student, student–instructor, and student–content interaction).

Recently, advances in Generative Artificial Intelligence (GenAI) (e.g., OpenAI's ChatGPT) have opened new possibilities for enhancing online interactivity (Kasneci et al., 2023). Large Language Models (LLMs), such as GPT-4 (Achiam et al., 2023) have demonstrated remarkable abilities in understanding and generating text, quickly establishing themselves as foundational components of many AI-powered applications, particularly in education domain. This technology supported the initiative of LLMapps, which refer to the type of software that uses LLMs as one of its building blocks, have been developed (Yan et al., 2024) in different downstream domains. Downstream refers to application-specific tasks where pre-trained LLMs are used or fine-tuned. For example, using GPT-4 to build an app that helps students write essays. Particulary in LLM-assisted apps in education like, Khanmigo, a GPT-4-based virtual tutor and classroom aide by Khan Academy to help student in their learning process (Anand, 2023).

The LLM-assisted apps can engage learners in dialogue, answer questions, and provide tailored feedback, potentially stimulating student engagement even in the absence of direct human instructors. While such AI tutors and facilitators are increasingly being integrated into educational settings (Kasneci et al. 2023), their effectiveness in low-resource contexts like Afghanistan remains under-studied.

Our work addresses these gaps by introducing an LLM-based conversational agent into online class discussions for Afghan students. We investigate whether a GPT-based powered chatbot, acting as a discussion facilitator alongside a human instructor, can enhance student engagement and improve learning outcomes. We conducted a randomized controlled trial to evaluate the chatbot's impact. We hypothesized that AI-facilitated discussions would lead to higher student participation (more frequent and longer contributions) and greater learning outcome experience compared to discussions led only by a human tutor. The study's focus on Afghan learners (including female students facing educational barriers) provides insight into AI's role in supporting equitable access to interactive learning.

The structure of the paper is as follows: Section 2 presents relevant studies. In Section 3, we detail the methodology, experimental settings, sample and sampling, and implementation of our social experiment. Section 4 presents the results, followed by discussions. Finally, we summarize our work and highlight future directions in section 5.

2. Related Work

Prior research emphasizes that a lack of interaction in virtual classrooms leads to diminished student engagement (Haleem et al., 2022). Another study found that larger online class sizes correlate with lower instructor performance in fostering interaction (Sorensen, 2015). Russell and Curtis, reported that students and teachers perceive less meaningful engagement in large-scale online language courses than in smaller ones (Russell & Curtis, 2013). Espasa and Meneses highlighted the importance of feedback processes in online environments, linking interactive feedback to better learning outcomes (Espasa & Meneses, 2010). Overall, ensuring interactivity in online education is critical to prevent passive learning during online educational course.

To address engagement challenges, various technology-mediated solutions have been explored. Deslauriers et al. demonstrated that actively engaging students in class (e.g., through interactive activities) yields greater actual learning, even if students subjectively feel they learned less (Deslauriers et al., 2019)—underscoring the importance of interactive engagement. Chandra and Palvia discussed the emerging paradigm of peer-to-peer learning in online education, suggesting that enabling students to interact with each other is key to improving outcomes (Chandra & Palvia, 2021). One promising approach to facilitate such interaction in online interaction is the use of conversational AI agents as a facilitator to support online activities (Ito et al., 2022).

Generative artificial intelligence (AI) is poised to revolutionize how humans work, and has already demonstrated promise in significantly improving human productivity (Bastani et al., 2024). LLMs, such as GPT-4 have demonstrated remarkable abilities in understanding and generating text, quickly establishing themselves as foundational components of many AI-powered applications, particularly in education domain (Achiam et al., 2023). For example, Bastani et al., conducted a field experiment with nearly 1,000 high school students to assess the impact of GPT-4 on math learning. They compared two AI tutors: GPT Base (standard ChatGPT interface) and GPT Tutor (prompt-engineered for learning support), integrated into 15% of the curriculum. While both improved short-term performance (48% and 127%, respectively), the study also found that students who later lost access to the AI performed worse than those who never used it, suggesting potential long-term dependency risks (Bastani et al., 2024).

Our study builds on this work from conservative lenses by applying GPT-4 as a conversational agent in an educational context to support downstream domains such as education. In line with this, we use GPT-based chatbot as facilitator to enhnace student enagement and learning outcome in online education. We rely on mainstream LLMs like GPT-4 as foundational model and tailored for specific post-lecture facilitation task. GPT-4 is a much-improved model compared to GPT-3.5, which is the large language model that powers the now infamous ChatGPT (Achiam et al., 2023).

Our work inspired by prior research on the efficacy of conversational agent in enhancing online discussion performance. Prior research on human-conversational agents behaviour studies suggests conversational agents can improve user interactions, enhance discussion and support minorities in online discussion. For example, Ito et al. developed an agent to facilitate crowd discussions in large groups (Ito et al., 2022). Conversational agents have been shown to increase women's contributions in online debates (Hadfi et al., 2023) and to improve overall engagement and problem-solving in discussion activities (Sahab et al., 2024). Argumentative chatbots can scaffold debates in online forums (Hadfi et al. 2021), and AI facilitators have demonstrated benefits for critical thinking in social media discussions (Tanprasert et al., 2024).

These studies indicate the potential of AI-driven agents to enrich online learning interactions. However, it is unclear how such tools perform in unique online education environments like Afghanistan, where resource limitations differ from typical settings (Haqbeen et al, 2023) can work. Our work addresses this gap by introducing an LLM-based conversational agent into online class discussions for Afghan students to empirically study how a Large Language Model (LLM)-based chatbot influences student engagement and learning outcomes in online courses in a low-resource educational setting like Afghanistan.

3. Methodology

3.1 Research design

The study recruited 80 undergraduate computer science students (40 female, 40 male; aged 18–26) from a pool of 3,432 qualified applicants (out of 3,761 who initially registered) from Afghanistan via local subjects' recruitment agency (Sahab et al., 2024b). A priori power analysis was conducted using G*Power v3.1 (Faul et al., 2009) to assess the adequacy of the sample size for an independent samples t-test comparing two groups (treatment vs. control). With a total of 80 participants (40 per group), the study is adequately powered (power ≈ 0.80) to detect large effect sizes (Cohen's $d \geq 0.65$) at a significance level of $\alpha = .05$.

We conducted a randomized controlled trial (RCT), and participants were randomly assigned to either a control or treatment condition (Sahab et al., 2024b). Each condition had 40 students during online lectures via Zoom, and subdivided into four discussion groups of 10 students each during post-lecture discussion via Discourse. All participants attended identical 50-minute online lectures followed by 40-minute post-lecture discussions moderated by a human instructor, and completed a follow-up self-report questionnaire. The treatment group additionally engaged in AI-facilitated discussions using a GPT-4-based chatbot during post-lecture discussion.

Figure 1 illustrates the recruitment and grouping process. Selected students were randomly assigned into two conditions (control vs. treatment, 40 each). To facilitate post-lecture discussions, each condition was further divided into four smaller discussion groups of 10 students (labeled A1–A4 for control and B1–B4 for treatment). We used stratified random sampling to ensure each discussion group had an equal gender mix (5 female, 5 male per group). All participants provided informed consent and then took part in study.

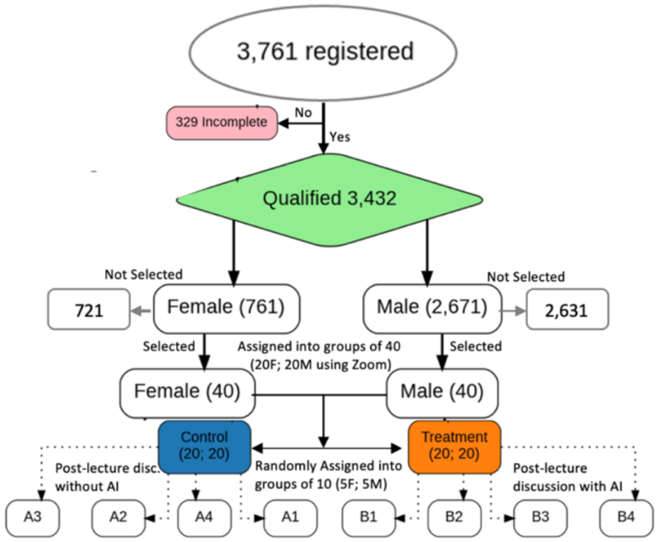

Figure 1. Provides an illustration of the recruitment process and the experimental pipeline.

3.1 Online lecture with follow-up discussion

At the beginning of experiment, both the control and treatment groups (n=80) attended an identical 50 minutes live lecture via Zoom, a videotelephone application which is used for online learning courses. The lecture content was the same for all students and was delivered by an experienced instructor (a Ph.D. in CS with >2 years online teaching experience). To maintain consistency across conditions, the lecture topic and materials were fixed: a lesson on "The Role of Information in Society." The instructor used the same slide presentation and script on both days to minimize any variability in instruction. Each lecture lasted approximately 50 minutes. By controlling all instructional content and delivery, we ensured that any differences observed between conditions could be attributed to the discussion facilitation method (with vs. without the AI chatbot) rather than differences in teaching.

Following the lecture, students were divided into groups of ten (5 female, 5 male per group), resulting in four discussion groups per condition (treatment and control) and engaged in a 40 minutes post-lecture discussion session conducted via an online customized discussion forum called Hyper-Democracy platform, which is built on the open-source Discourse system (Ito et al., 2024; Discourse, 2025). Each group's discussion space was supervised randomly by the same course instructor, who served as the human facilitator across both conditions.

In each condition group (*n* = 40), a single instructor was assigned to facilitate both conduct the Zoom lectures and interact (moderate) the subsequent post-lecture discussions. These discussions were conducted in smaller breakout groups, with 10 students per group across four subgroups. The instructor rotated between these groups at random intervals. This setup was designed to reflect the practical challenges instructors face in large-scale online courses, particularly the difficulty of independently managing and stimulating meaningful discussions across multiple simultaneous groups.

In the treatment condition, an AI chatbot facilitator was integrated alongside the instructor to support and stimulate the conversation. In contrast, the control condition relied solely on human facilitation by the instructor. This setup ensured that all participants took part in an online discussion following the lecture, with the key distinction being the presence or absence of AI facilitation.

The discussion session was structured to prompt reflective engagement around the guiding question: "What are the key takeaways from today's lecture?" To facilitate meaningful post-lecture discussion, the session was divided into three clearly defined phases:

1. Key Point Generation (15 minutes): Participants were encouraged—by the human facilitator (instructor) and/or the AI agent—to identify and articulate the most significant concepts presented in the lecture.

2. Bridging Theory with Practice (15 minutes): Participants were encouraged—by the facilitator and/or the AI agent—to connect lecture content with real-world contexts or personal experiences.

3. Summary and Forward Reflection (10 minutes): Participants were encouraged—by the facilitator and/or the AI agent—to summarize the discussion, extract key insights, and offer feedback or suggestions for future topics.

This structured format was designed to ensure consistency across conditions and to allow systematic analysis of the AI facilitator's impact. The overall experimental procedure is summarized in Figure 2.

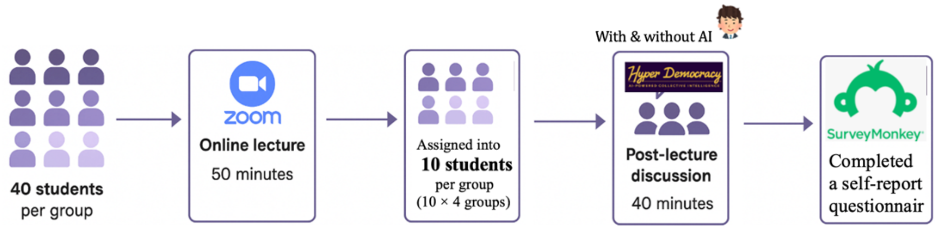

Figure 2. The main experimental pipeline.

3.2 Chatbot Design and Deployment

The conversational agent used in the treatment condition was implemented using the OpenAI's GPT-4 (Achiam et al., 2023). To ensure the chatbot's interventions were relevant and pedagogically useful, we employed a few-shot learning and prompting approach (Tunstall et al., 2022) during its development. We used the GPT-4 model behavior by providing several example question-answer pairs related to the lecture content and discussion guidelines. To ensure both coherence and contextual appropriateness in agent responses, we implemented a predefined prompt adaptation and learning strategy (Liu et al., 2023). Instead of prompting GPT to generate sentences from scratch, we curated a structured set of facilitation messages (a set of 17 human-authored facilitation messages) in advance for post-lecture discussion session while adopting three clearly defined phases explained in section 3.1. During interactions, the agent leverages GPT to post exact facilitated messages which pre-defined by human author or dynamically modify these predefined sentences to better align with the ongoing discussion. This indicates that the chatbot does not always generate messages autonomously; rather, it draws on previously authored human facilitation posts as references. In Figure 3, red-bordered box shows the chatbot's posted previously written human-authored facilitated message, and blue-bordered box a conversation between a student and the chatbot, where chatbot posted a facilitated message entirely created on its own based on posted examples. This design balances the need for consistency in facilitation objectives with the flexibility required to respond to context-specific dialogue. In essence, GPT is not tasked with creating new content but with adapting existing prompts to fit the evolving conversational context. We adjusted the phases for chabtot facilitation described in section 3.1 based on time.

To integrate user (students) interactions with AI agent during post-lecture discussion, we deployed the chatbot into the Discourse (Ito et al., 2024), so the participant could post messages and replies in the discussion threads. On the backend, the chatbot listened to the discussion in real time via the forum's API and decided to intervene once it sees posted messages. On the frontend, the chatbot's profile name was explicitly set as "Facilitator", and the chatbot's profile picture was set similar to those from the popular Japanese "Irasutoya" site (Irastutoya, 2025), which is often used for free educational or presentation illustrations.

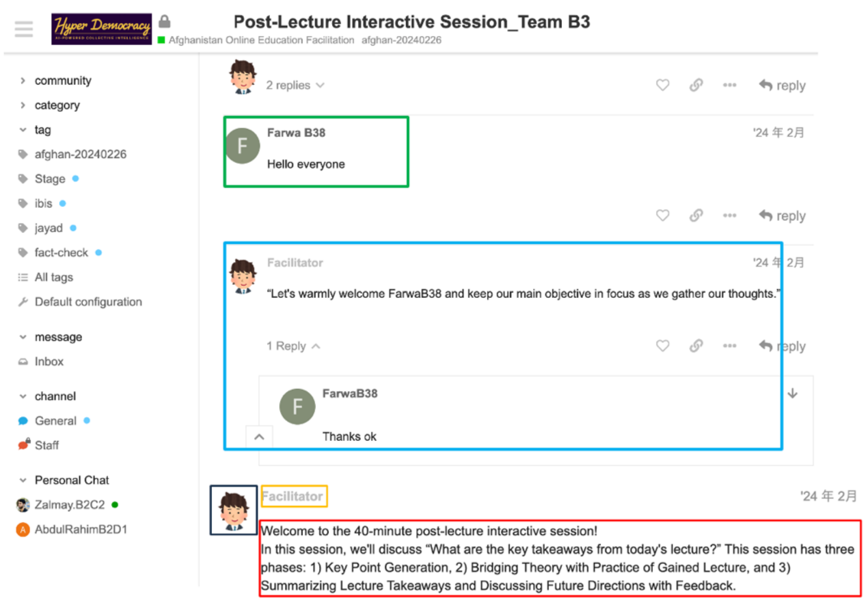

Figure 3. The main user interface of the post-lecture discussion website featuring AI-facilitated interaction (Treatment).

In Figure 3, the interface illustrates a structured online discussion session conducted with treatment group. The green-bordered box at the top shows a message posted by a human participant. Below, the blue-bordered section represents a human–chatbot interaction, in which the AI facilitator acknowledges the participant and encourages focus. On the left, the dark blue-circled avatar depicts a cartoon-style profile image similar to those from the popular Japanese "Irasutoya" site representing the AI facilitator. The orange-highlighted label indicates the chatbot's displayed role "Facilitator". Finally, the red-bordered box shows the chatbot's posted human-authored facilitated message.

3.3 Data collection and analysis

We logged all discussion activity on the forum for both conditions. Each student's posts, replies, word-count, and likes were recorded, along with chatbot facilitated messages (treatment groups), timestamps and thread structure. To evaluate student-student engagement and discussion performance, we defined quantitative metrics based on prior research on online discussions (Hadfi et al., 2023; Sahab et al., 2024a): (1) Posts per user – the average number of messages each student posted; (2) Replies per user – the average number of direct replies to others' posts; (3) Word count – the total number of words contributed by each student, and length of posted messages; and (4) Likes received – the number of "likes" each student's posts received from peers. Among these indicators, the number of posts, replies, and total word count reflect active participation and the depth of discussion, whereas likes primarily signal peer appreciation (Gao, 2016; Schreiner & Fischer, 2021) and are therefore considered light contributions. Metrics were computed to observe any changes or learning effects.

In addition to log data, we administered a short post-experiment survey at the end. This survey assessed students' self-reported learning outcomes and perceived confidence in the subject matter. In particular, we asked students to rate their agreement with the statement: "As a result of this class setting, I feel more confident in this area of study." on a 5-point Likert scale (1=Strongly Disagree, 5=Strongly Agree). We used the 5-point Likert scale to determine the attitudes of individuals towards acceptance study setting. The Likert scale, developed by Rensis Likert in 1932, is a reliable and valid method for measuring attitudes (Sullivan & Artino, 2013). The aim was to capture whether the students felt the discussion format (with or without AI) improved their confidence and perceived learning outcome in the subject.

For analysis, we compared the control and treatment groups on the engagement metrics and survey results. We used independent-samples t-tests to assess differences between conditions for each metric, as well as for the self-report ratings. Statistical significance was set at $p < 0.05$. All quantitative analyses were performed using IBM SPSS Statistics (v28). Below, in section 4, we report the results.

4. Results

4.1 post-lecture discussion engagement

Figure 4 summarizes the post-lecture discussion performance metrics for the control vs. treatment groups.

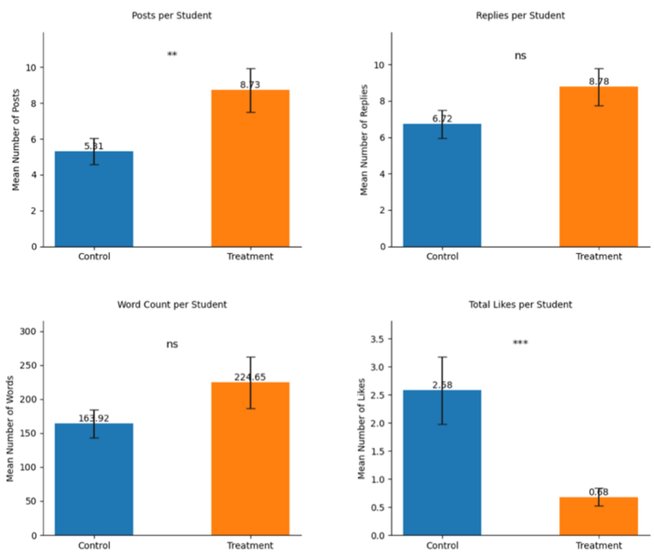

Figure 4. Discussion engagement metrics between control (n=36) and treatment (n=40) groups.

Our finding suggests that treatment group posted significantly more messages per student than the control (p<0.01), while differences in replies and word count were not significant (ns). The control group received more likes per student (*p<0.001). Treatment group students made an average of 8.73 posts each, significantly exceeding the control group's 5.11 posts per student (t test, p < 0.01). This indicates that the chatbot's presence stimulated students to contribute more frequently to the discussion. For the other engagement metrics, the treatment group demonstrated a slightly higher mean number of replies per student (8.78) compared to the control group (6.72), as well as a greater average word count per student (224.65 vs. 67.92). However, these differences were not statistically significant. An interesting finding was that control group posts received significantly more "likes" from peers on average than treatment group posts (2.88 vs. 0.78 likes per student, p < 0.001).

4.2 Self-reported learning outcome

To assess students' perceptions of their learning experience, we compared the post-experiment survey results between the two conditions. Figure 5 presents the distribution of responses for the control and treatment groups. Overall, students who participated in the AI-facilitated discussions reported higher confidence in their understanding of the course material. The treatment group's average rating was higher than the control groups. An independent t-test confirmed that this difference was statistically significant (t(78) = –2.125, p = 0.037).

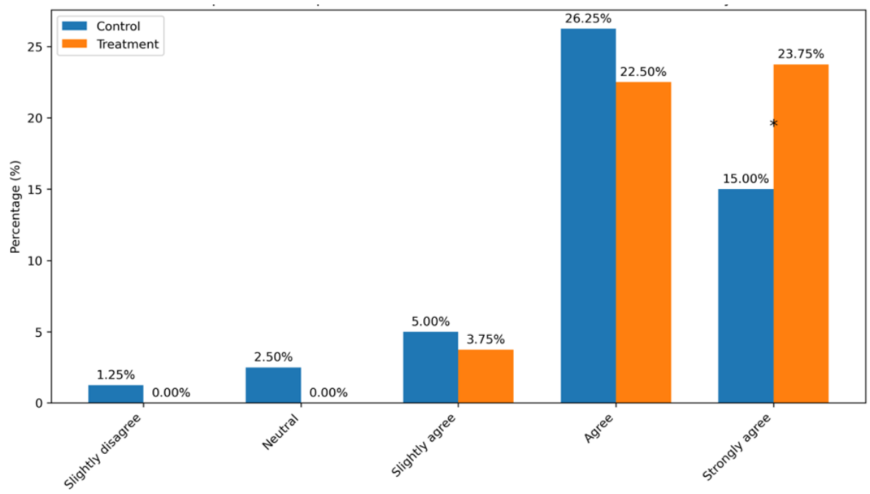

Figure 5. Comparison of post-intervention self-report measurement responses of users between control (n=40) and treatment (n=40) groups.

5. Discussion and Conclusion

In this study, we investigated the impact of an GPT-based AI conversational agent on student engagement and learning outcome in an online course interactive session in Afghanistan. Using a controlled experimental design, we found that embedding a GPT-4-based chatbot facilitator into post-lecture discussions significantly improved multiple facets of student engagement. Analysis of discussion logs and post-intervention surveys revealed that the treatment group (discussion with the AI facilitator) demonstrated significantly higher participation rates, with more posts and replies, during post-lecture discussion and reported greater confidence in their understanding of the course material.

An interesting observation emerged regarding the number of "likes" received across conditions. Posts in the control group (discussion without the AI facilitator) received more likes from peers than those in the treatment group. One possible explanation is that the human-only discussions in the control group resulted in fewer overall posts, allowing students to focus their attention—and consequently their likes—on a smaller number of contributions. In contrast, the higher volume of posts in the treatment groups, facilitated by the chatbot, may have diluted peer attention across more content. This interpretation aligns with the *bandwagon effect* (Schmitt-Beck, 2015), wherein individuals are more likely to endorse content that has already received visible support. In the control condition, where engagement was relatively lower, students may have shown their support through likes rather than additional contributions (*posts* & *replies*). Conversely, in the treatment groups, the chatbot's active facilitation may have encouraged more students to participate directly by posting (*posts* & *replies*), reducing their reliance on passive indicators of engagement such as *likes*.

In this study, our primary focus was on active participation and content engagement, as these measures provide a more meaningful indication of learning involvement (Gao, 2016; Schreiner & Fischer, 2021). Specifically, we emphasized students' efforts to articulate their opinions by posting, rather than merely expressing feelings or offering surface-level reactions by liking. In summary, these results confirm that the chatbot had a strong positive effect on student engagement in the post-lecture discussions.

These finding are further supported by self-reported questionnaire data, which indicated that students who participated in chatbot-facilitated discussions reported higher levels of confidence in the subject matter compared to those in the human-only condition. The higher self-reported confidence in the treatment group on subject matters also merits AI role as a facilitator in discussion. Students in the AI-supported discussions felt more confident about the study material, which is consistent with research showing that interactive and responsive learning environments improve student self-confidence and satisfaction (Pence, 2022). Chatbot likely contributed by promptly addressing confusion and keeping the discussion productive, thus reinforcing students' understanding. Taken together, the survey results reinforce the behavioral data (discussion logs), highlighting the value of AI-assisted facilitation in enhancing student engagement and perceived learning outcomes.

Our findings are in line with emerging evidence that facilitating peer learning can enhance educational experiences (Chandra & Palvia, 2021), interactive and responsive learning environments improve student self-confidence and satisfaction (Pence, 2022), and AI facilitators can amplify the voices and opinions in discussions (Hadfi et al., 2023, Sahab et al., 2024b), and overall support engagement opportunities (Sahab et al., 2024a).

Despite these findings, the study has several limitations and should be addressed in future research. First, while the sample size in our study is adequately powered (power ≈ 0.80) to detect large effect sizes (Cohen's $d \geq 0.65$) at a significance level of $\alpha = .05$. However, it is underpowered to detect medium ($d = 0.50$) and small ($d = 0.20$) effects, for which the estimated power falls below the conventional threshold of 0.80. This limitation should be considered when interpreting non-significant results. Additionally, the study was limited to a single-day intervention with computer science undergraduates considering one environment. Future research should consider replicating the experiment over multiple sessions across different educational programs and areas to examine the consistency and generalizability of the observed effects. Another limitaiton of this study is the chatbot was configured to support multiple facets of engagement: instructor–student interaction (by answering questions or providing prompts), student–student interaction (by encouraging students to respond to each other), and student–content interaction (by bringing in relevant points from the lecture). However, this study focused exclusively on student–student interactions. Future research should explore additional dimensions of interactivity, such as student–instructor and student–content interactions, to assess the consistency and generalizability of the observed effects.

In conclusion, our study provides empirical evidence that AI-powered conversational agents can play a supportive role to enhance student-student interactions in online education. The GPT-based chatbot in our study functioned effectively as a facilitator during post-lecture discussion. These results are among the first to demonstrate such benefits in an Afghan educational setting, suggesting the approach holds promise for other resource-constrained environments. While our study focused on student-to-student interactions, future work will explore the impact of AI facilitation on student–instructor and student–content interactions. We believe integrating AI agents across these dimensions could further optimize online learning experiences. An extended version of this research, including qualitative analysis of discussion content, is in preparation for submission to the IEICE Transactions on Information and Systems.

Acknowledgement

- **Funding or Grant:** This research was partially supported by the JSPS KAKENHI (Grant Number: 23K17164, Japan) and JST CREST (Grant Number: JP- MJCR20D1, Japan) funds.
- **Data/Software Access Statement**: The discussion logs data is available upon reasonable request at https://hyper-democracy.com/tag/afghan-20240226 (accessed date 21 April 2025); and data from the questionnaire are available upon request from the corresponding author.
- **Contributor Statement**: Conceptualization, J.H. and S.S.; methodology, J.H., S.S.; software, J.H. and T.I.; validation, J.H. and S.S.; formal analysis, J.H. and S.S.; investigation, J.H. and S.S.; resources, J.H., S.S. and T.I.; data curation, J.H. and S.S.; writing—original draft preparation, J.H. and S.S.; writing—review and editing, J.H. and S.S.; visualization, J.H. and S.S. supervision, T.I.; project administration, J.H. and S.S.; funding acquisition, J.H. and T.I. All authors have read and agreed to the published version of the manuscript.
- **Conflict Of Interest (COI)**: The authors declare no conflict of interest.

References

Hart, C. M., Friedmann, E., & Hill, M. (2018). Online course-taking and student outcomes in California community colleges. *Education Finance and Policy, 13*(1), 42–71.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys, 55*(9), 1–35. https://doi.org/10.1145/3505244

Salmon, A. K., & Barrera, M. X. (2021). Intentional questioning to promote thinking and learning. *Thinking Skills and Creativity, 40*. https://doi.org/10.1016/j.tsc.2021.100822

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. (2023). *GPT-4 technical report*. arXiv. https://arxiv.org/abs/2303.08774

Yan, C., Ren, R., Meng, M. H., Wan, L., Ooi, T. Y., & Bai, G. (2024). Exploring ChatGPT app ecosystem: Distribution, deployment, and security. In *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering* (pp. 1370–1382).

Anand, P. (2023, March 16). Khan Academy creates GPT-4 based helper Khanmigo marking formal entry of AI into education. *Dataquest*.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). Generative AI can harm learning. *Research Paper, The Wharton School*. https://doi.org/10.2139/ssrn.4895486

Haqbeen, J., Sahab, S., Ito, T., & Rizzi, P. (2021). Using decision support system to enable crowd identify neighborhood issues and its solutions for policy makers: An online experiment at Kabul municipal level. *Sustainability*, 13(10), 5453. https://doi.org/10.3390/su13105453

Szopiński, T., & Bachnik, K. (2022). Student evaluation of online learning during the COVID-19 pandemic. *Technological Forecasting and Social Change, 174*, 121195.

Haqbeen, J., Sahab, S., & Ito, T. (2023). A digital initiative to address girls' education challenges in collaboration with NPO in post-2021 Afghanistan. *IIAI Letters on Information and Interdisciplinary Research, 4*(1).

Lyons, E. (2023). Household and student survey-based evidence on girls' secondary education in Afghanistan. *Social Sciences & Humanities Open, 8*(1), 100533.

Shayan, Z. (2015). Gender inequality in education in Afghanistan: Access and barriers. *Open Journal of Philosophy, 5*, 277–284.

Singh, A. K., & Meena, M. K. (2022). Challenges of virtual classroom during COVID-19 pandemic: An empirical analysis of Indian higher education. *International Journal of Evaluation and Research in Education, 11*(1), 207–212.

Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. S. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers, 3*, 275–285.

Sorensen, C. (2015). An examination of the relationship between online class size and instructor performance. *Journal of Educators Online, 12*(1), 140–159.

Russell, V., & Curtis, W. (2013). Comparing a large- and small-scale online language course: An examination of teacher and learner perceptions. *The Internet and Higher Education, 16*, 1–13.

Espasa, A., & Meneses, J. (2010). Analyzing feedback processes in an online teaching and learning environment: An exploratory study. *Higher Education, 59*(3), 277–292.

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences, 116*(39), 19251–19257.

Chandra, S., & Palvia, S. (2021). Online education next wave: Peer to peer learning. *Journal of Information Technology Case & Application Research, 23*(3), 157–172.

Kasneci, E., Sessler, K., Winkel, K., Schmid, U., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274.

Ito, T., Hadfi, R., & Suzuki, S. (2022). An agent that facilitates crowd discussion. *Group Decision and Negotiation, 31*, 621–647.

Hadfi, R., Okuhara, S., Haqbeen, J., et al. (2023). Conversational agents enhance women's contribution in online debates. *Scientific Reports, 13*, 14534.

Sahab, S., Haqbeen, J., & Ito, T. (2024a). Conversational AI as a facilitator improves participant engagement and problem-solving in online discussion. *IEICE Transactions, E107-D*(4).

Hadfi, R., et al. (2021). Argumentative conversational agents for online discussions. *Journal of Systems Science and Systems Engineering, 30*, 450–464.

Tanprasert, T., Fels, S., Sinnamon, L., & Yoon, D. (2024). Debate chatbots to facilitate critical thinking on YouTube: Social identity and conversational style make a difference. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA.

Sahab, S., Haqbeen, J., Hadfi, R., et al. (2024b). E-contact facilitated by conversational agents reduces interethnic prejudice and anxiety in Afghanistan. *Communications Psychology, 2*, 22.

Hyper-Democracy. (2025). *Hyper-Democracy Platform*. https://hyper-democracy.com

Discourse. (2025). *Discourse Platform*. https://www.discourse.org

Ito, T., Dong, Y., Haqbeen, J., Matsuo, T., & Sahab, S. (2024). Multiple AI agents that support crowd discussion. In *Proceedings of the 12th ACM Collective Intelligence Conference*, Northeastern University, Boston, MA.

Sullivan, G. M., & Artino, A. R., Jr. (2013). Analyzing and interpreting data from likert-type scales. *Journal of Graduate Medical Education*, 5(4), 541–542. PubMed. https://doi.org/10.4300/JGME-5-4-18

Tunstall, L., et al. (2022). Efficient few-shot learning without prompts. *arXiv preprint*, arXiv:2209.11055.

Tofade, T., Elsner, J., & Haines, S. T. (2013). Best practice strategies for effective use of questions as a teaching tool. *American Journal of Pharmaceutical Education, 77*(7), 155.

Schmitt-Beck, R. (2015). Bandwagon effect. In *The International Encyclopedia of Political Communication* (pp. 1–5).

Gao, L. (2016). Motives for posting and liking on social media: Comparing differences between Chinese and European users (Master's thesis). University of Twente. https://essay.utwente.nl/69110/

Schreiner, M., Fischer, T., & Riedl, R. (2021). Impact of content characteristics and emotion on behavioral engagement in social media: Literature review and research agenda. *Electronic Commerce Research, 21*, 329–345.

Pence, P. L. (2022). Student satisfaction and self-confidence in learning with virtual simulations. *Teaching and Learning in Nursing, 17*(1), 31–35.

Irasutoya. (2025). *2025 年*. https://www.irasutoya.com/2025/