Generative AI as Personal Tutor? Designing a Study Planner Bot to Facilitate Online Creative Computing Education

1. Project Overview

To recap, the study is situated within the context of enhancing students' learning experience in distance/online education. Previous studies show that a high level of self-regulated learning (SRL) plays a significant role in shaping students' learning progression and ultimately leads to successful completion of this type of education (Dumford and Miller, 2018; Schwarzenberg and Navón, 2020). Inspired by previous studies on utilizing chatbot for SLR (Du, Huang and Hew, 2021; Hew et al., 2023), this project further investigates the use of chatbot technology as a technological intervention to improve self-regulation among online students, with a specific focus on creative technology education.

The primary objective is to assist students in developing personalised and suitable study plans. To achieve this, the study adopts the design science methodology (Hevner and Storey, 2021) and conducts two rounds of design iterations. The final design is a study planner bot that is integrated into a popular social application—Discord—and is powered by a large language model—GPT-3.5—which serves as the primary mechanism for personalized recommendations. Following prior research on the SMART framework (i.e., specific, measurable, assignable, realistic, and time-related) for educational purpose (Aghera *et al.*, 2018), the AI chatbot takes users' input as training data and develops achievable learning objectives based on this framework.

For the usability study, the research recruited seven participants who are fully engaged in online learning and registered at two UK universities, to evaluate the performance and adoption of the study planner bot in an experimental setting. Overall, the findings showed that participants had a positive experience interacting with the bot, as measured by *perceived ease of use*, *usefulness*, and *enjoyment of interaction*. The results also demonstrated that users appreciated integrating such an assistive educational technology into a social platform for convenient use. Looking forward, the study proposes that considering a sense of trust is essential for future studies when developing AI study bots with recommendation capabilities.

2. Research Methods and Materials

Chatbot Design & Research: Overview

Drawing from experimental HCI research, this study reviews several case studies that involved the design and implementation of chatbots for educational purposes and employs the Design Science research methodology, focusing on four aspects: resources, governance, operational intervention, and knowledge bases, to develop the study planner bot project (Jain *et al.*, 2018; Hevner and Storey, 2021; Khalil and Rambech, 2022; Hew *et al.*, 2023). In practice, the study is carried out in two stages. Firstly, it applied the waterfall methodology to create the study planner bot. While several educational chatbot design projects, such as Khalil and Rambech (2022), advocate for the application of the agile method due to its flexibility in the technological development process, this study opted for a different approach. Considering the project's complexity and confined timeframe, the linear approach of the waterfall model was deemed better suited to ensure a clear path for chatbot development and to effectively achieve the initial research objectives. Although user interaction during each phrase of the waterfall methodology is minimal, it remains crucial for finalizing a successful product. To maximise the value of user feedback and inform the sequential design process, one participant was invited to contribute during the chatbot validation phase.

In the second stage—usability testing, the research involves measuring user experience by inviting potential end users (i.e., online education students) to assess and evaluate the bot's utility and usability. Participants will be asked to complete a brief questionnaire, which measures their previous experiences with chatbot technologies and explores four aspects: *perceived ease of use*, *usefulness*, *interactional engagement*, and *sense of trust*. Suggested in previous chatbot research, these measures can help better understand the bot's potential use and influence on improving self-regulation skills such as formulating achieveable study objectives for online education (Casas *et al.*, 2020; Du, Huang and Hew, 2021; Hew *et al.*, 2023). Additionally, the survey includes two open-ended questions asking for end users' suggestions for future improvements.

Chatbot Configuration: Discord API

Discord is a free social networking platform mainly for group and private communication. Previous studies such as Vladoiu and Constantinescu (2020) established an online computer science education community utilizing Discord. Their decision of selecting Discord as the host lies in its capacity to facilitate users' simultaneous presence and connections across multiple communities. This is complemented by Discord's user-friendly interface and data infrastructure, which allows novel users to easily create and manage their own communities (Vladoiu and Constantinescu, 2020). By the time of writing, the University of the Arts London (UAL) has already set up several community servers on Discord, including the one for the Creative

Computing Institute (CCI). Given its ease of use and the significant presence of UAL students on the Discord, this study designed and integrated the bot within the Discord platform.

Discord provides developers with a programming interface known as the Discord API[1]. For this project, the researcher has chosen to utilize the discord.py[2] library, which is written in Python. This library contains pre-defined functions that enable communication between end users and the Discord server.

Chatbot Configuration: OpenAI API

The core part of the chatbot is powered by the OpenAI's latest generative large language model GPT-3.5[3], which has been specifically optimized for chat applications. With pre-determined prompting questions, the generative model can help the chatbot process user inputs and generate study planning recommendations. The initial iteration of the chatbot development was created in Python within the Google Colab environment[4] (see Figure 01). During the early prototyping phase, the study primarily focused on empirically evaluating the performance of OpenAI's language model via its API and understanding the structure of the API to integrate it with the Discord API to develop the bot. For instance, Figure 01 illustrates how to initiate a conversation using the generative language model by specifying three pre-defined roles, namely, *system*, *user*, and *assistant* to stimulate a natural language conversation between the generative model and user. As a result, these self-initiated experiments helped to decide on the most suitable model (i.e., gpt-3.5-turbo) for the study's purpose and confirm the initial prompts to input into the model, thereby guiding study recommendation creation.

[1] https://discord.com/developers/docs/intro
[2] https://discordpy.readthedocs.io/en/latest/
[3] https://platform.openai.com/docs/models
[4] https://colab.research.google.com/drive/1VyFxdsvk2hDn_IpBGmdmmXUbl8vcBNHf?usp=sharing

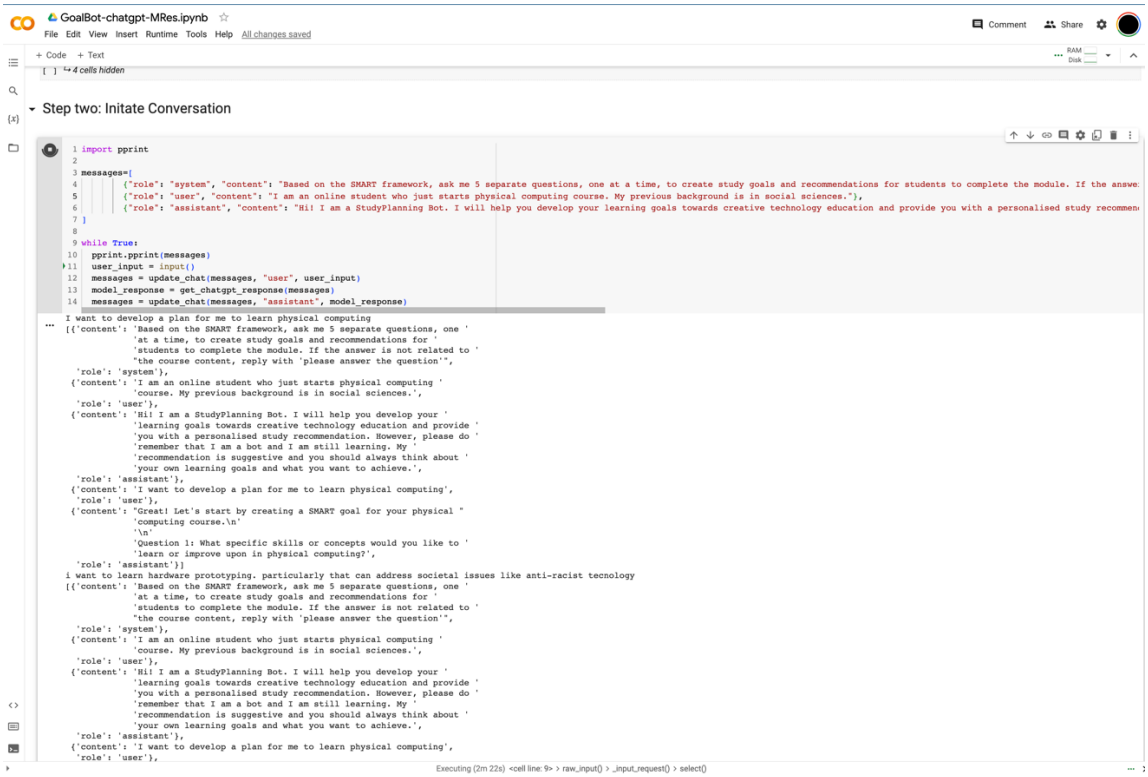

Fig 01. A screenshot of experimenting OpenAI API

User Experience Research

A total of seven participants, aged between 21 and 32 years old, were recruited for the usability testing in the final stage, while one participant was involved in the validation stage of chatbot development. Participants were recruited through convenience sampling (e.g., personal contacts) and snowball methods. All seven participants were registered as online/distance learning students in the field of creative technology at two UK universities. In the sense, the experiment was conducted entirely online via Microsoft Teams. Participants were encouraged to use their assigned ID as their screen name during each session (see figure 02 for more details).

In each session, the researcher began by introducing the participant information, research goals, and experimental procedures. After receiving the participants' verbal consent, they were instructed to use Discord login details to access the bot. Participants were then asked to share their screen activities with the researcher and utilize the Think-aloud method to describe their intentions and behaviors. Each experiment lasted 10 minutes without any interference. After this test play, participants were asked to complete a short anonymous survey[5] dispatched through Microsoft Forms for product evaluation.

[5] https://forms.office.com/e/uhLCHGvdEv

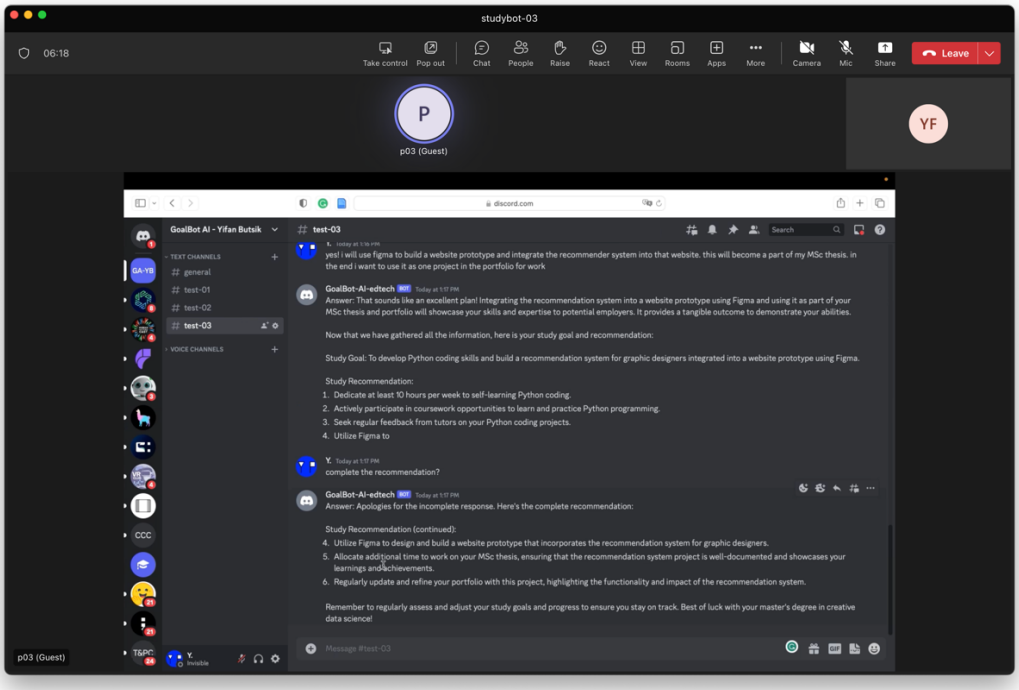

Fig 02. A screenshot of online usability testing with participant 03

3. Results and Discussions

Q1: How can a study planner chatbot be designed to facilitate self-regulated learning in online higher education?

Following the Design Science research methodology, the study planner chatbot underwent two rounds of development before being presented to the interested research participants. The first iteration primarily focused on modifying pre-determined prompts, which assisted the language model in initiating the conversation for enquiring about the user's background information. As previously mentioned, the bot is powered by OpenAI's GPT-3.5-turbo language model. Figure 03 demonstrates the function of initiating the generative model. The initial prompt message (as shown in Figure 03) is configured to enable the language model processing personalized study plans based on the SMART framework for setting learning goals (Bjerke and Renger, 2017; Aghera *et al.*, 2018), while contextualizing these plans in the creative technology education. As a result, the prompt used to clarify the model's assistant role (see Figure 03, line 26, "initial_prompt_message") needs to emphasize keywords such as "SMART framework" and enquiry styles like "ask questions one at the time". Together, these can enhance acquiring knowledge about user needs through a more interactive user experience.

```
23      def start_conversation(self, user_message: str) -> str:
24          self.messages = []
25
26          self.messages.append({"role": "assistant", "content": initial_prompt_message})
27          self.messages.append({"role": "user", "content": user_message})
28          self.messages.append({"role": "system", "content": initial_response})
29
30          return initial_response
```

Fig 03. A screenshot of code script for initiating the GPT model

The second iteration focused on chatbot validation. A participant was invited to assist in refining the bot's functionality. During this phase, two aspects of the chatbot design were improved: (a) simplification of the start commands, and (b) incorporation of a warning message into the greetings. According to Discord's developer documentation[6], start commands (or "slash commands") are a new feature for interacting with bots on the Discord platform. Initially, to guide novice users in starting a conversation with the study planner bot, intuitive and common commands such as "*/ai*", "*/chatter*", and "*/chatbot*" were implemented. However, after consulting with the participant, it was found that these commands were redundant. The researcher then decided to adopt the participant's suggestion of using more specific commands like "*/edubot*" and "*/studybot*", which better highlight the bot's purpose and distinguish it from other chatbot usage on the Discord server.

Another enhancement involved the chatbot's initial message delivered to the user (see Figure 03, line 28, "initial_response"). Researchers have assessed the abilities and capabilities of the GPT model for recommendation tasks and concluded that while ChatGPT excels at forecasting, it doesn't perform as effectively in direct recommendation tasks (Liu *et al.*, 2023). Following these findings, the chatbot's initial response was designed to include a warning message, informing the user about the suggestive nature of the bot's recommendations. This approach also served to enhance user trust. Below the Figure 04 presents selected snapshots of final design of the study planning bot after the two iterations of chatbot development.

[6] https://support.discord.com/hc/en-us/articles/1500000368501-Slash-Commands-FAQ

Fig 04. Selected screenshots of final design (p03 usability testing)

Q2: How do students experience and perceive a conversational agent for study goal planning integrated within social applications?

Seven participants (2 male, 4 female, 1 non-binary) completed the user experience questionnaire. To address potential bias from their previous chatbot interactions, it's important to note that although all participants had experience with chatbot technology prior to the experiment, none had utilized a chatbot for educational purposes. Figure 05 illustrates participants' previous use cases: travel is the most common industry and banking and shopping are positioned in the second place. It is interesting to note that chatbot use also is also incorporated into governmental services, as indicated in the "other" category.

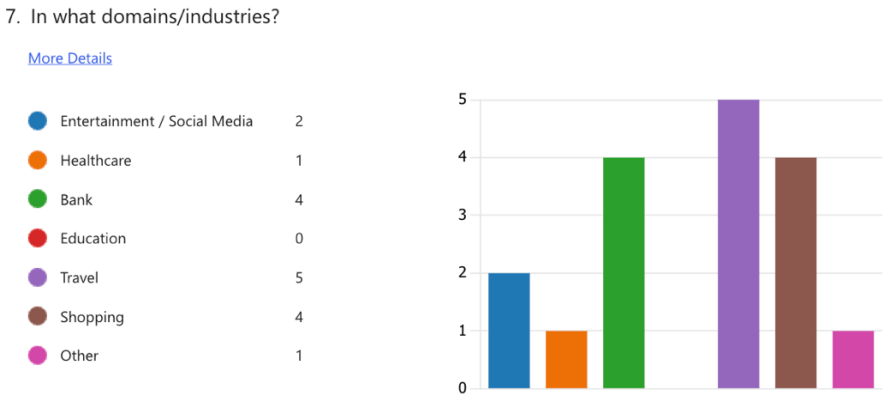

Fig 05. Chatbot use cases

The study bot evaluation focuses on four dimensions: *usefulness*, *ease of use*, *engagement*, and *trustworthiness*. Participants were requested to reflect on their interaction experience and rate these aspects using a 5-point scale (1=low, 5=high). Below the table 01 to 04 presents an overview of the participants' ratings. The first two measurements, namely usefulness and perceived ease of use, are suggested in the Technology Acceptance Model (TAM) to measure the possibility of adopting new technologies (Davis,1989). As reflected in Tables 01 and 02, the perceived usefulness (M=3.11, SD=0.69) of the study planner bot is slightly higher than its ease of use (M=3.07, SD=0.85). That said, Table 01 indicates that participants gave the bot's ability to enhance study planning skills a low average rating of 2.43. This might be due to the fact that the assessment was conducted through a brief, 10-minute experiment, and skill enhancement generally requires a longer period to take effect. Overall, participants reported that the bot could potentially have a positive impact on study planning actions in online computing education. Notably, participants found the bot particularly useful for refining and reflecting on their past study goals (M=3.71, SD=0.49). However, when it comes to functionality, participants encountered difficulties (M=2.71, SD=0.76) and issues with handling error messages (M=2.86, SD=1.07), which may increase their reluctance to use such technology.

Enjoyment of engagement and sense of trust are two measurement scales frequently used to evaluate AI-based conversational agents (Borsci et al., 2022). In this study, these two scales are used to understand whether the study planner bot design can deliver an engaging and trustworthy user experience. The results, as shown in Tables 03 and 04, suggest that participants found the bot to offer a higher level of interactive engagement (M=3.18, SD=0.66) than trustworthiness (M=2.64, SD=0.68). Notably, the bot was praised for its timely responses. However, when it came to the aspect of trust, participants felt that both the interaction style of the bot and the content of its recommendations were not entirely trustworthy. One potential explanation for this could be that the bot's recommendation process was not transparent and failed to gain user trust in the customized study goal plans. Another potential reason could be a mismatch between users' expectations and the bot's actual affordance, as the recommendations are powered by a language model-GPT that has yet to be fine-tuned. This discrepancy may have resulted in lower trust ratings, underscoring the importance of aligning user expectations with the system's capabilities.

Participants were asked to answer two open-ended questions as part of the chatbot experience evaluation. Regarding the first question "*Will you utilize the study planning bot to facilitate your online study? Why (not)?*", five out of seven participants responded positively. They appreciated the innovative service and the bot's availability regardless of time and location. The negative feedback pointed to the trustworthiness of the information recommended by the bot, which aligns with the findings from the previous descriptive

statistical reports. The second question focused on evaluating the integration of the chatbot into a social application. All participants reported favorably and expressed interest in the design and use of the study planner bot on Discord owning to its convenient use, "reminder" role, and shareability for other student groups. However, one participant suggested moving the bot to a more work-oriented platform such as Slack to maintain the distinction between social and work/study life.

Table 01. Summary of survey data: usefulness (1=low, 5=high)

	Item	N	Mean	SD
Usefulness	I found the bot was effective to create a personalized study goal for me.	7	3.43	.79
	I found the bot enabled me to reflect on my study plan and goals.	7	3.71	.49
	I found the bot enhanced my study planning and goal settings skills.	7	2.43	.79
	I found study planning was easier by communicating with the bot.	7	2.86	.69

Table 02. Summary of survey data: perceived ease to use (1=low, 5=high)

	Item	N	Mean	SD
Easiness	I found it easy to initiate the conversation.	7	2.71	.76
	I found it easy to communicate my thoughts to the bot.	7	3.29	.76
	I found the bot responded/behaved under my expectation.	7	3.43	.79
	I found it easy to handle error messages encountered in the interaction.	7	2.86	1.07

Table 03. Summary of survey data: interactional engagement (1=low, 5=high)

	Item	N	Mean	SD
Engagement	I found the study planning process through the bot was fun and enjoyable.	7	3.14	.69
	I found conversing with the bot was exciting.	7	2.71	.76
	The bot was able to engage with my enquires in an acceptable timeframe.	7	3.86	.38
	The bot made it fun to provide the study plan information.	7	3.00	.82

Table 04. Summary of survey data: sense of trust (1=low, 5=high)

	Item	N	Mean	SD
Trustworthiness	The bot tried to understand my needs for study planning.	7	3.29	.76
	I found the bot provided a reliable study recommendation plan.	7	2.71	.76
	I developed trust in what the bot told me.	7	2.71	.49
	The bot reassured me that I can trust this technology.	7	1.86	.69

4. Conclusion & Future Work

Centred on enhancing self-regulation skills for online creative computing education, the project presents a study planner bot as a technological intervention. Built upon several earlier studies of designing and implementing educational bot for higher education (Du, Huang and Hew, 2021; Hew *et al.*, 2023), this project furthers exploration and contribution to the field in two ways: First, it develops and embeds the bot within a popular social networking platform—Discord and utilizes one of the latest generative machine learning models—GPT, to power personalized study recommendations. Second, it broadens the measurement scales by incorporating "sense of trust" and "interactional engagement" to better understand potential adoption of such technology. The research findings suggest that participants overall had a positive experience with the study planning bot integrated into a social platform. Participants particularly noted that the bot could prompt further reflection on their goal-setting process on a daily basis, serving as a "reminder" within their daily communication app. However, the study also uncovered that trust is a crucial factor for participants when considering the adoption of the chatbot for their online learning. The current recommendation systems did not provide relevant features that could enhance user trust.

There are several limitations exist that affect the interpretation of the research findings. Firstly, due to the specificity of the research topic (i.e., online creative technology education), the sample size of the user experience study is relatively small (N=7). Secondly, this study examines participants' chatbot experiences by measuring immediate effects during a short, 10-minute user testing session. It remains uncertain whether users' perceptions of the bot would change if they were to interact with it over a full academic term. Thirdly, since the bot was developed using the Discord API, the user experience is significantly influenced by Discord's inherent affordances (e.g., user interface and interaction design).

For future work, three potential improvements are proposed. Firstly, it might be beneficial to utilize objective evaluation methods such as analyzing users' behavioral engagements and utterance turns as complementary approaches to the self-reported style of measurement. Secondly, as suggested in the participants' survey feedback, the study could experiment with various large language models beyond the GPT family. Constructing a unique recommender system based on fine-tuned generative algorithms may yield beneficial results. Lastly, this research indicated that user trust is a significant factor influencing the study bot experience. Future work could address this issue by adopting concepts such as explainable AI to present the recommendation-building process to the user, thereby enhancing potential adoption of this technology in their self-regulated learning process.

5. Reference

Aghera, A. et al. (2018) 'A Randomized Trial of SMART Goal Enhanced Debriefing after Simulation to Promote Educational Actions', Western Journal of Emergency Medicine, 19(1), pp. 112–120. Available at: https://doi.org/10.5811/westjem.2017.11.36524.

Bjerke, M.B. and Renger, R. (2017) 'Being smart about writing SMART objectives', Evaluation and Program Planning, 61, pp. 125–127. Available at: https://doi.org/10.1016/j.evalprogplan.2016.12.009.

Borsci, S. et al. (2022) 'The Chatbot Usability Scale: the Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents', Personal and Ubiquitous Computing, 26(1), pp. 95–119. Available at: https://doi.org/10.1007/s00779-021-01582-9.

Casas, J. et al. (2020) 'Trends & Methods in Chatbot Evaluation', in Companion Publication of the 2020 International Conference on Multimodal Interaction. ICMI '20: INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION, Virtual Event Netherlands: ACM, pp. 280–286. Available at: https://doi.org/10.1145/3395035.3425319.

Davis, F.D. (1989) 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology', MIS Quarterly, 13(3), p. 319. Available at: https://doi.org/10.2307/249008.

Du, J., Huang, W. and Hew, K.F. (2021) 'Supporting students goal setting process using chatbot: implementation in a fully online course', in 2021 IEEE International Conference on Engineering, Technology & Education (TALE). 2021 IEEE International Conference on Engineering, Technology & Education (TALE), pp. 35–41. Available at: https://doi.org/10.1109/TALE52509.2021.9678564.

Dumford, A.D. and Miller, A.L. (2018) 'Online learning in higher education: exploring advantages and disadvantages for engagement', Journal of Computing in Higher Education, 30(3), pp. 452–465. Available at: https://doi.org/10.1007/s12528-018-9179-z.

Hevner, A.R. and Storey, V.C. (2021) 'Externalities of Design Science Research: Preparation for Project Success', in L. Chandra Kruse, S. Seidel, and G.I. Hausvik (eds) The Next Wave of Sociotechnical Design. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 118–130. Available at: https://doi.org/10.1007/978-3-030-82405-1_14.

Hew, K.F. et al. (2023) 'Using chatbots to support student goal setting and social presence in fully online activities: learner engagement and perceptions', Journal of Computing in Higher Education, 35(1), pp. 40–68. Available at: https://doi.org/10.1007/s12528-022-09338-x.

Jain, M. et al. (2018) 'Evaluating and Informing the Design of Chatbots', in Proceedings of the 2018 Designing Interactive Systems Conference. New York, NY, USA: Association for

Computing Machinery (DIS '18), pp. 895–906. Available at: https://doi.org/10.1145/3196709.3196735.

Khalil, M. and Rambech, M. (2022) 'Eduino: A Telegram Learning-Based Platform and Chatbot in Higher Education', in P. Zaphiris and A. Ioannou (eds) Learning and Collaboration Technologies. Novel Technological Environments. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 188–204. Available at: https://doi.org/10.1007/978-3-031-05675-8_15.

Liu, J. et al. (2023) 'Is ChatGPT a Good Recommender? A Preliminary Study'. arXiv. Available at: https://doi.org/10.48550/arXiv.2304.10149.

Schwarzenberg, P. and Navón, J. (2020) 'Supporting goal setting in flipped classes', Interactive Learning Environments, 28(6), pp. 671–684. Available at: https://doi.org/10.1080/10494820.2019.1707691.

Vladoiu, M. and Constantinescu, Z. (2020) 'Learning During COVID-19 Pandemic: Online Education Community, Based on Discord', in 2020 19th RoEduNet Conference: Networking in Education and Research (RoEduNet). 2020 19th RoEduNet Conference: Networking in Education and Research (RoEduNet), pp. 1–6. Available at: https://doi.org/10.1109/RoEduNet51892.2020.9324863.