



AI-Powered Chatbot Assistant for Course Organization

Alessio Magnolfi^(✉) , Eugenio Vicario , and Leonardo Boncinelli 

Università degli Studi di Firenze, 50127 Florence, Italy
{alessio.magnolfi,eugenio.vicario,leonardo.boncinelli}@unifi.it

Abstract. In this paper we present the design and evaluation of an AI-powered chatbot tool to support course-related information for an undergraduate class at the University of Florence. Using the Landbot platform integrated with OpenAI's gpt4 model, the chatbot provided students with real-time responses to course-related frequent questions. We assessed the chatbot's performance through a post-course survey administered to 35 students. The ease of use, speed, and reliability were highly appreciated, while usefulness and response adequacy received mixed feedback. The reported main limitation was its limited scope, only answering questions about few selected topics. A k-means clustering analysis revealed two distinct groups of students, with a different satisfaction level; however, both groups agreed on preferring to use the chatbot over the official website of the University of Florence.

Keywords: Generative Artificial Intelligence · Course Assistant · Syllabus · Higher Education · User Experience

1 Introduction

Chatbots are computer software systems designed to interact in a natural conversation style with humans, to facilitate a wide range of tasks. With the aid of Natural Language Processing they have been improved in simulating a human conversational style and improving customer service.

The impact of Artificial Intelligence is becoming more and more important on a variety of sectors, such as in business related activities [1], healthcare [2], and finance [3]. Following the strong development and adoption of chatbots in many fields, recent research has been conducted on the study of human-agent interactions in conversational settings [4–8].

The educational setting is among the various fields where the aid of AI has enabled significant improvements and also transformed existing practices and ways of working [9]. Conversational agents found their way into schools and universities enabling a scaling up of good teaching practices [10] and allowing an equal access to learning material to all students [11]. This type of application is spreading very rapidly, and many institutions are implementing various technological solutions. At the same time, the scientific literature is developing with the aim of evaluating the effectiveness of chatbots in higher education [12, 13].

Thanks to latest advancements in this technology, educational field practices could reach big steps in offering new kinds of services and improve the existing ones. Chatbots are used to assist, mentor or aid learning and they are often a supplement to the already existing educational practices [14–16]. Students perceive AI chatbots as useful in getting feedback or help with writing, coding, and academic tasks. Though they are also concerned about their accuracy and reliability and about potential negative impacts on their learning processes, critical thinking, discipline, and creativity [17]. A meta-analysis performed by Wu & Yu [18] showed that AI chatbots have a large effect on learning outcome and that this impact is bigger in higher education in comparison to primary and secondary education. Though, this effect tends to diminish in long interventions, indicating that the novelty of the tool could mediate such effects.

Chatbots are useful not only for knowledge related activities: another important aspect of AI tools applied to higher education settings is about streamlining repetitive and time-consuming tasks that occur both in administration offices and for university teachers. They can participate in activities such as scheduling, answering queries about school policies, and assisting with enrollment processes [19]. For example, thanks to automated grading systems, university teachers can get rid of the most repetitive tasks and concentrate more on their core activities like designing lessons and personalizing feedback [20,21].

Chan et al. [22] built a chatbot for providing course information and students' opinion, analysis of possible course choices and providing recommendations. The chatbot performed better in comparison to the program leader and peers, but it performed worse in comparison to the website and the programme booklet. Though, the number of participants in this study was very low.

Alaa and Abdelhafez [23] developed a chatbot to assist new students and their parents into the university admission process. They created the chatbot with Botsify platform and tested it on forty-two students. Then they tested it with a confusion matrix and the Chatbot Usability Questionnaire (CUQ).

Hien et al. [24] investigated the most requested tasks for an administrative chatbot in higher education: information related to courses, facilities and timetables were the most common ones. Above 80% of the more than one hundred participants stated that they think a chatbot is useful for the following tasks: providing content of an exercise and notifying activities deadline; supporting course content and learning resources research, providing information on scholarships and graduation requirements, schedules, facilities and courses. Moreover, all tasks surveyed have reached a minimum useful approval by 40% of the students, indicating that conversational agents are very promising in this sector.

Georgia State University (GSU) introduced a virtual assistant named Pounce [25] to tackle obstacles to enrollment faced by students transitioning from high school to college. Given their high number of people that do not show up during fall enrollment, they tried to tackle common obstacles directly through an automated chatbot. After the first year of implementation, students who used

the tool were more likely to submit a final transcript, attend orientation, accept a college loan and complete college loan counseling.

Our research explores the development and implementation of an educational chatbot designed for an undergraduate program at the University of Florence. The design of the tool will be presented in Sect. 2. To assess the impact we adopted a mixed approach using a questionnaire with both quantitative questions and open feedbacks regarding the user experience of the chatbot, as detailed in Sect. 3. Results of the study will be then presented in Sect. 4, while in Sect. 5 potential implications and future studies will be discussed.

2 Implementation of the Chatbot

We developed a chatbot specifically for the *Economia Politica* (Economics) undergraduate course offered by the Economics Department at the University of Florence, with the primary aim of assisting students in accessing course information in an efficient way.

The chatbot was designed using the *Landbot* platform. We decided to use such a platform mainly for its user-friendly interface and the possibility to share the chatbot directly through a link. *Landbot* also allows integration with other tools like *Google Sheet* and the possibility to implement both rule based and AI based interactions. *Google Sheets* has been used to check matriculation numbers and also to store every question made by students and the relative answer the chatbot provided.

Rule based interactions have been used for greetings, the first check for matriculation numbers and a pre-made answer when the chatbot could not retrieve information from its corpus of text. All other interactions happened through *OpenAI* APIs, using *GPT-4* model.

To lower costs, we decided not to insert the whole text entirely, but we broke it down into two phases:

- Recognition of the argument related to the question;
- Prompting the argument related corpus text for question answering.

The type of information of the corpus has been divided into the five following arguments: general information about the course, books and material used, course program, didactic calendar and exams information.

After onboarding the user, each question was categorized by *GPT-4* (via APIs) into one of the predefined areas. Then, only the relevant information corpus related to the specific topic was processed again through *GPT-4* APIs to generate a response to the user's question. This allowed us to cut the costs of the model, reaching an average of 1 cent of dollar per each pair of question and answer.

The chatbot was available through the entire course and it could provide real-time responses with a 24/7 availability. The main user interface pages are shown in Appendix A.

3 Method

We conducted an observational study for practical reasons and with an exploratory purpose, potentially leading to future experimental investigations. Additionally, an observational approach was more viable for ethical reasons, as we could not exclude any students from the opportunity to use the tool.

The chatbot has been presented at the first day of lecture, at the end of February 2024. Students followed the course for the entire semester and then, at the conclusion of the course, we conducted a survey to evaluate the effectiveness and ease of use of the chatbot.

The survey included both quantitative metrics in a 7 point Likert scale, and three open questions for personal feedback, allowing students to express their personal experiences and suggestions for improvement.

Likert questions are about eight important aspects of the user experience over a chatbot:

- Ease of use: “the chatbot has been easy to use”;
- Usefulness: “the chatbot has been useful”;
- Interactivity: “the chatbot has been interactive”;
- Expectations: “the chatbot lived up to my expectations”;
- Reliability: “the chatbot has been reliable”;
- Question understanding: “the chatbot could often understand my questions”;
- Answer adequacy: “the chatbot could often answer in an appropriate way”;
- Speed: “the chatbot could often answer quickly”;
- Preference: “I prefer to use the chatbot over the official Unifi website”.

After completing the questionnaire, means have been analyzed and a k-means clustering has been conducted in order to check if people could be grouped in two or more user types.

The survey also comprehended three open questions regarding the positive and negative aspects of the AI tool. The questions are the following:

- “What’s the best aspect of the chatbot?”;
- “What’s a big limitation instead?”;
- “Do you want to leave a free comment? Each feedback will be appreciated and will help us improving the service”.

Successively, a manual review has been conducted in order to group each answer into a broader category. Each answer has been read by authors and then grouped into one common aspect. If an answer could not be grouped into any cluster, it has been discarded. To allow anonymity, we did not ask participants any identification number when filling the questionnaire, hence we were not able to track specific questions to questionnaire results.

4 Results

4.1 Descriptive Statistics: Closed Questions

A total of 133 completed interactions occurred with the chatbot and 35 students participated in the post-course survey. All results are shown in Table 1, while a detailed distribution of answers can be seen in Figs. 1, 2, and 3.

Table 1. Summary of descriptive statistics for survey’s closed questions. Questions are shown in Method section.

	easy	useful	interactive	expectations	reliable	understanding	answering	speed	preference
mean	5.31	4.17	4.69	4.74	5.06	4.46	4.91	5.11	4.94
std	1.75	1.65	1.71	1.60	1.57	1.92	1.72	1.79	1.41

L'utilizzo del chatbot è stato, a mio parere (risposte da 1 a 7):

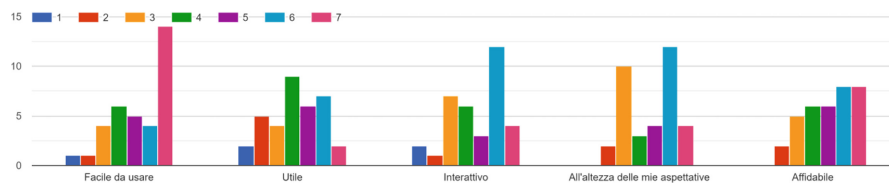


Fig. 1. First set of questions: the chatbot has been easy to use; the chatbot has been useful; the chatbot has been interactive; the chatbot lived up to my expectations; the chatbot has been reliable.

From the results of the survey we can clearly see that students appreciated the use of the chatbot while there are still few areas of improvement. Overall, the chatbot has been useful and it provided a quick and reliable way to answer students’ queries about the course information. All answers of the questionnaire have a mean greater than 4 in a scale out of 7. The chatbot was very easy to use (mean = 5.31), could answer quickly (mean = 5.11) and it was reliable (mean = 5.06). However, the perceived usefulness was not high (mean = 4.17). Expectations (mean = 4.74), level of understanding (mean = 4.46) and adequacy in answers (mean = 4.91) had good overall values but also great variance, with the majority of answers being either 3, 6 or 7. This could entail a different approach from different users, with some of them capable of gaining the most out of such technology while others are more disappointed by it. Finally, students preferred the chatbot over the traditional way of getting information through the social website (mean = 4.94).

4.2 Descriptive Statistics: Open Questions

Regarding the open feedback, results are very clear: speed and ease of use has been cited openly by the majority of students (18 and 8 respectively) as the best

Il sistema riusciva spesso a (risposte da 1 a 7):

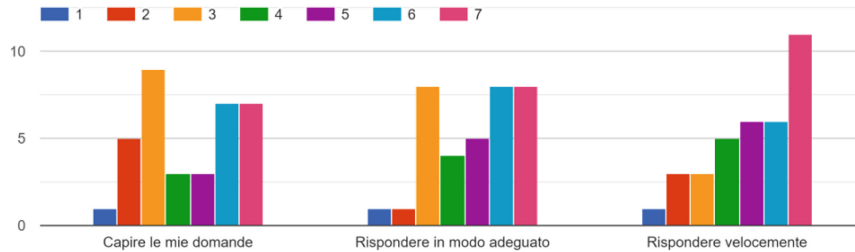


Fig. 2. Second set of questions: the chatbot could often understand my questions; the chatbot could often answer in an appropriate way; the chatbot could often answer quickly.

Preferisco usare il chatbot anziché cercare le informazioni sul sito ufficiale Unifi
35 risposte

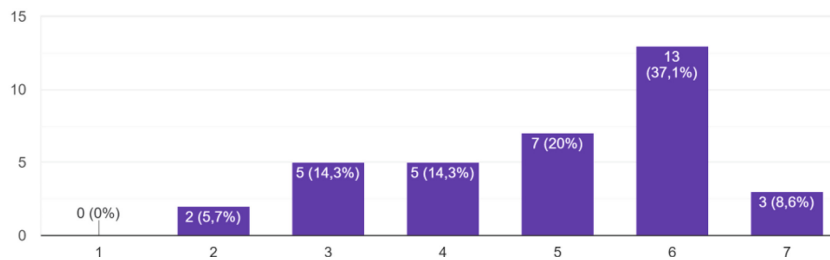


Fig. 3. Preference of using the chatbot over the official University of Florence website.

aspect of the chatbot. However, most of the students (21 out of 35) pointed out that the responses were limited to a few issues of the course.

The most common positive categories, according to the students, are the following:

- Speed (mentions of speed, quickness, or immediacy in response): 20 answers out of 35;
- Convenience (ease of use and practicality): 8 answers;
- Information quality and quantity (providing good, useful or sufficient information): 4 answers.

The worst aspect cited by students are the following:

- Limited Scope (only answers specific questions, limited to specific topics): 12 answers;
- Understanding (ability to understand questions): 5 answers.

4.3 Cluster Analysis

The results of the cluster analysis on student responses regarding their interaction with the chatbot suggest a clear distinction between two groups. Participants have been grouped into n clusters, ranging from 1 to 10, such that the Sum of Squared Errors (SSE) is minimized for each value of n . SSE has been calculated as the squared distance of each participant from the group mean across all 8 questions. The initial SSE with one group is 315, while considering two groups the sum of squared errors already drops to 157, see Fig. 4. Adding a third group lowers the SSE to 141: a very minimal change. Beyond two clusters, the reduction in SSE is generally much more gradual, supporting the decision to focus on the division into two primary groups. We will refer to such clusters with the names of “C1” and “C2”. Cluster 1 represents students who provided consistently higher average scores across all evaluated dimensions of chatbot performance, while Cluster 2 represents students with lower average scores. Statistical analysis (Table 2) confirms that the differences between these groups are not only large but also statistically significant, as indicated by the p-values being consistently below 0.01 for all questions, except for the item “preference”, which shows no significant difference between the groups.

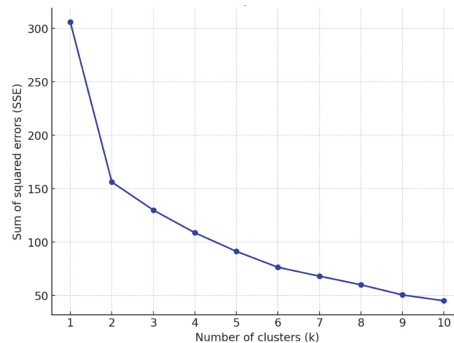


Fig. 4. Elbow method for optimal k-clustering.

The effect sizes (Cohen-d) regarding all items highlight the importance of these differences, with all values but one exceeding 1 in absolute values and several values exceeding 2, signifying a strong practical significance. For instance, the effect size of 3.51 for users’ expectations indicates a large difference in how people perceived the chatbot. This big difference implies that group “C1” members had their expectations met by the chatbot to a much greater extent than those in group “C2”. Other items, such as “usefulness”, “interactivity” and “reliability” also show large effect sizes (2.75, 2.65 and 2.46 respectively), suggesting that these aspects of the chatbot have been perceived particularly differently

Table 2. T-test between Cluster 1 (C1) and Cluster 2 (C2) for each question. Details regarding questions are shown in Method section.

Question	C1 Mean (std)	C2 Mean (std)	Difference	Cohen-d
easy	6.19 (1.47)	4.00 (1.24)	-2.19***	-1.61
useful	5.24 (1.00)	2.57 (1.02)	-2.67***	-2.65
interactive	5.81 (0.93)	3.00 (1.11)	-2.81***	-2.75
expectations	5.86 (0.85)	3.07 (0.73)	-2.79***	-3.51
reliable	6.05 (0.92)	3.57 (1.09)	-2.48***	-2.46
understanding	5.57 (1.54)	2.79 (0.97)	-2.79***	-2.17
answering	5.81 (1.17)	3.57 (1.55)	-2.24***	-1.63
speed	6.14 (1.06)	3.57 (1.55)	-2.57***	-1.93
preference	5.00 (1.45)	4.86 (1.41)	-0.14	-0.10

from student to student. The other four aspects of the user experience, namely “ease of use”, “question understanding”, “answer adequacy” and “speed” had a lower but still big size effect.

The only question that did not differ statistically from the two groups was the preference to use the chatbot over the official Unifi website. This result suggests that students’ preferences are not strongly influenced by individual characteristics. The overall average rating is 4.94, which is above the neutral point of 4.00 on the Likert scale, indicating a generally positive level of satisfaction with the chatbot.

Cluster analysis is not sufficient to give a clear answer to such issues, and further studies are needed to have a clearer understanding how chatbot might be perceived useful and if there is a specific subgroup of users who can benefit more from such digital tools.

5 Discussion

The implementation of the chatbot to assist students on information related to a university course has been perceived as very reliable, fast and easy to use by students, who had generally a positive experience of it. Its limited functioning to a single course has however been perceived as a limitation of its usefulness and effectiveness. Students might be willing to use such a technology to a broader extent, including also other university courses or more general information about the university activities or bureaucracy.

The strong results in k-means clustering suggests that students' experiences can be categorized into two groups: students who rate positively the chatbot as a valuable, reliable and easy-to-use tool and students who do not perceive these qualities as much. These differences could be influenced by various factors, such as individual preferences for technology, prior exposure to similar tools, or specific expectations regarding chatbot interactions. Further studies could take into account the Unified Theory of Acceptance and Use of Technology (UTAUT) model [26] or the AI Device Use and Acceptance (AIDUA) model [27] in order to check if the propensity to use technology tools in general or to use AI devices in particular is correlated with the cluster of students who perceived the chatbot more positively. Also, the motivation in using AI products might play a role in shaping interactions with the chatbot: the Questionnaire of Artificial Intelligence Use Motives (QAIUM) [28] could as well be inserted in the battery of tests for future works.

The absence of a significant difference in the preference for the chatbot over the official website might indicate that while students perceive differently its functionality and performance, they agree on preferring the chatbot over the official website to get information related to the course syllabus.

The small number of participants is a big limitation of the study, being only 35 students who participated in it. Further studies with larger number of participants are needed to confirm our results.

Moreover, the impact of the chatbot has been studied in isolation, without comparisons to other means of obtaining information, such as the official website or other tools. It would be advisable to conduct both a pre- and post-evaluation of users' satisfaction with information gathering from different sources, such as websites and peers, while conducting only a post-evaluation of their satisfaction with gathering information through the chatbot.

Implementing an objective evaluation of the experience (response time, accuracy..) was beyond our scope of the study, since we interested in the UX rather than in validating the chatbot from a technical point of view. Though adding such metrics alongside the subjective reports could yield insights on how user perceptions might depend also on the performative aspects of the chatbot.

By addressing these limitations with future studies, a more comprehensive understanding of students' engagement with AI-driven chatbots could be achieved, improving their effectiveness as information retrieval tool inside the academic environment.

A Appendix

(See Figs. 5, 6 and 7).



Fig. 5. First onboarding message, greeting the student. This part of the chatbot has been designed through a rule based approach.



Fig. 6. Matriculation check, in order to restrict the use of the chatbot only to "Economia Politica" students. This part of the chatbot has been designed through a rule based approach.

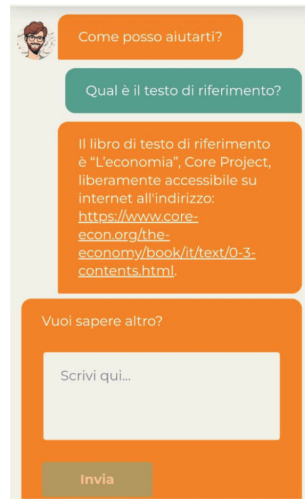


Fig. 7. Core part of the chatbot, where students' questions are answered in a conversational way. This part of the chatbot has been designed through *GPT-4* APIs.

References

1. Enholm, I.M., Papagiannidis, E., Mikalef, P., Krogstie, J.: Artificial intelligence and business value: a literature review. *Inf. Syst. Front.* **24**(5), 1709–1734 (2022)
2. Singhal, K., et al.: Large language models encode clinical knowledge. *Nature* **620**(7972), 172–180 (2023)
3. Mhlanga, D.: Industry 4.0 in finance: the impact of artificial intelligence (AI) on digital financial inclusion. *Int. J. Financ. Stud.* **8**(3), 45 (2020)
4. Stein, N., Brooks, K., et al.: A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. *JMIR Diab.* **2**(2), e8590 (2017)
5. Numata, T., et al.: Achieving affective human-virtual agent communication by enabling virtual agents to imitate positive expressions. *Sci. Rep.* **10**(1), 5977 (2020)
6. Altay, S., Hacquin, A.-S., Chevallier, C., Mercier, H.: Information delivered by a chatbot has a positive impact on COVID-19 vaccines attitudes and intentions. *J. Exp. Psychol. Appl.* **29**(1), 52 (2023)
7. Bilancini, E., Boncinelli, L., Vicario, E.: AI-powered chatbots: effective communication styles for sustainable development goals. *arXiv preprint [arXiv:2407.01057](https://arxiv.org/abs/2407.01057)* (2024)
8. Cai, N., Gao, S., Yan, J.: How the communication style of chatbots influences consumers' satisfaction, trust, and engagement in the context of service failure. *Humanit. Soc. Sci. Commun.* **11**(1), 1–11 (2024)
9. Al-Zahrani, A.M., Alasmari, T.M.: Exploring the impact of artificial intelligence on higher education: the dynamics of ethical, social, and educational implications. *Humanit. Soc. Sci. Commun.* **11**(1), 1–12 (2024)
10. Ferguson, R., Sharples, M.: Innovative pedagogy at massive scale: teaching and learning in MOOCs. In: *Open Learning and Teaching in Educational Communities: 9th European Conference on Technology Enhanced Learning, EC-TEL 2014, Graz, Austria, 16–19 September 2014, Proceedings 9*, pp. 98–111. Springer (2014)

11. Virtanen, M.A., Käärjäinen, M., Liikanen, E., Haavisto, E.: The comparison of students' satisfaction between ubiquitous and web-based learning environments. *Educ. Inf. Technol.* **22**, 2565–2581 (2017)
12. Neupane, S., et al.: From questions to insightful answers: building an informed chatbot for university resources. arXiv preprint [arXiv:2405.08120](https://arxiv.org/abs/2405.08120) (2024)
13. Mogavi, R.H., Deng, C., Kim, J.J., Zhou, P., Kwon, Y.D., Metwally, A., Tlili, A., Bassanelli, S., Bucchiarone, A., Gujar, S., et al.: ChatGPT in education: a blessing or a curse? a qualitative study exploring early adopters' utilization and perceptions. *Comput. Hum. Behav.: Artif. Hum.* **2**(1), 100027 (2024)
14. Nee, C.K., Rahman, M., Yahaya, N., Ibrahim, N.H., Razak, R.A., Sugino, C.: Exploring the trend and potential distribution of chatbot in education: a systematic review. *Int. J. Inf. Educ. Technol.* **13**(3), 516–525 (2023)
15. Labadze, L., Grigolia, M., Machaidze, L.: Role of AI chatbots in education: systematic literature review. *Int. J. Educ. Technol. High. Educ.* **20**(1), 56 (2023)
16. Pérez, J.Q., Daradoumis, T., Puig, J.: Rediscovering the use of chatbots in education: a systematic literature review. *Comput. Appl. Eng. Educ.* **28**(6), 1549–1565 (2020)
17. Schei, O.M., Møgelvang, A., Ludvigsen, K.: Perceptions and use of AI chatbots among students in higher education: a scoping review of empirical studies. *Educ. Sci.* **14**(8), 922 (2024)
18. Wu, R., Yu, Z.: Do AI chatbots improve students' learning outcomes? Evidence from a meta-analysis. *Br. J. Edu. Technol.* **55**(1), 10–33 (2024)
19. Lee, J., Suh, T., Roy, D., Baucus, M.: Emerging technology and business model innovation: the case of artificial intelligence. *J. Open Innov.: Technol. Mark. Complexity* **5**(3), 44 (2019)
20. Celik, I., Dindar, M., Muukkonen, H., Järvelä, S.: The promises and challenges of artificial intelligence for teachers: a systematic review of research. *TechTrends* **66**(4), 616–630 (2022)
21. Liu, B., Sundar, S.S.: Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychol. Behav. Soc. Netw.* **21**(10), 625–636 (2018)
22. Ho, C.C., Lee, H.L., Lo, W.K., Lui, K.F.A.: Developing a chatbot for college student programme advisement. In: 2018 International Symposium on Educational Technology (ISET), pp. 52–56. IEEE (2018)
23. Aloqayli, A., Abdelhafez, H.: Intelligent chatbot for admission in higher education. *Int. J. Inf. Educ. Technol.* **13**(9), 1348–1357 (2023)
24. Hien, H.T., Cuong, P.-N., Nam, L.N.H., Nhung, H.L.T.K., Thang, L.D.: Intelligent assistants in higher-education environments: the fit-eBot, a chatbot for administrative and learning support. In: Proceedings of the 9th International Symposium on Information and Communication Technology, pp. 69–76 (2018)
25. Page, L. C., Gehlbach, H.: How an artificially intelligent virtual assistant helps students navigate the road to college. *Aera Open* (2017)
26. Venkatesh, V., Morris, M. G., Davis, G. B., Davis, F. D.: User acceptance of information technology: toward a unified view. *MIS Q.* 425–478 (2003)
27. Gursoy, D., Chi, O.H., Lu, L., Nunkoo, R.: Consumers acceptance of artificially intelligent (AI) device use in service delivery. *Int. J. Inf. Manage.* **49**, 157–169 (2019)
28. Yurt, E., Kasarci, I.: A questionnaire of artificial intelligence use motives: a contribution to investigating the connection between AI and motivation. *Int. J. Technol. Educ.* **7**(2) (2024)