



Evaluating User Experience in an Educational Chatbot Creation Platform

Edirlei Soares de Lima^(✉)  and Margot M. E. Neggers^{}

Academy for AI, Games and Media, Breda University of Applied Sciences,
Breda, The Netherlands

{soaresdelima.e,neggers.m}@buas.nl

Abstract. While extensive research has evaluated chatbot usability, the user experience of chatbot creation platforms remains largely unexplored. This study addresses this gap through a diary-based study of the user experience evolution within an educational chatbot development system. Ten students from an Applied Data Science and Artificial Intelligence program participated in a two-week academic challenge requiring them to create compliance-focused chatbots using an internal chatbot creation platform. Data collection occurred at three time points through detailed diary entries combining qualitative reflections and quantitative usability measures adapted from Unified Theory of Acceptance and Use of Technology (UTAUT) framework. Results reveal consistent improvement trajectories across usability dimensions alongside qualitative evidence of skill development through a two-phase learning curve: initial exploration followed by competency consolidation. These findings provide evidence-based guidance for designing educational chatbot development platforms that support meaningful user learning trajectories.

Keywords: Chatbot Creation Platforms · User Experience · Longitudinal Study · Learning Trajectory · Large Language Models

1 Introduction

The integration of Artificial Intelligence (AI) technologies in higher education has led to increased interest in chatbot platforms as tools for enhancing learning experiences and educational services [4]. Modern chatbot platforms, powered by Large Language Models (LLMs), offer opportunities to deliver contextually relevant responses across various educational applications [20]. While commercial AI platforms provide sophisticated capabilities, educational institutions often require solutions that maintain institutional control over data and align with specific pedagogical and administrative needs [9]. This has driven the development of specialized chatbot creation platforms designed specifically for educational contexts [7], enabling educators, staff, and students to develop customized conversational agents without extensive technical expertise.

Despite the proliferation of chatbot creation tools and platforms, research on user experience with these development environments remains limited. While extensive research has evaluated the usability and effectiveness of chatbots themselves [15, 16], there is a notable lack of investigation into how users experience the platforms and tools used to create these chatbots. Existing studies of development platforms have primarily focused on technical comparisons and feature analyses rather than user experience evaluation. This gap is particularly significant given that the usability of creation tools directly impacts the quality of the resulting chatbots. Furthermore, most existing usability evaluations of educational technology platforms rely on single-session studies or short-term assessments, providing only snapshots of user experience rather than insights into how interactions, satisfaction, and skill development evolve during extended use these AI development tools.

This paper investigates user experience evolution in educational chatbot creation platforms through a longitudinal diary-based study. We examine how university students as platform users experience and adapt to the chatbot creation process, investigating how their perceptions, satisfaction, and interaction with the platform change over a two-week period of sustained use. Our research addresses two key questions: (1) How do users' perceptions and satisfaction with chatbot creation platforms evolve during extended use? and (2) What factors influence user experience trajectories and skill development in AI creation tools? Through an analysis of user diary entries collected at three time points, we provide insights into the temporal dynamics of user experience that can inform the design of more effective chatbot development platforms in educational contexts.

The paper is organized as follows. Section 2 reviews related work. Section 3 describes the study methodology and the chatbot development platform. Section 4 presents the results and discussion, and Sect. 5 concludes with key insights and implications for designing educational chatbot development platforms.

2 Related Work

Extensive research has focused on evaluating the usability and user experience of chatbots themselves. Studies have developed specialized evaluation instruments such as the Chatbot Usability Questionnaire (CUQ) [8] and the BOT Usability Scale (BUS-15) [2], recognizing that traditional usability measures may not adequately capture the unique aspects of conversational interfaces. These studies have consistently demonstrated that chatbots require different evaluation approaches due to their conversation-driven functionality and natural language interaction paradigms.

In contrast, research on the evaluation of chatbot development platforms remains limited. Most studies have primarily focused on technical comparisons rather than user experience. For example, Dagkoulis and Moussiades [6] conducted a comparative evaluation of major commercial chatbot development platforms (Google, Microsoft, Amazon, and IBM), defining technical criteria and

calculating scores based on requirement assumptions, but without examining user interaction or experience aspects. Similarly, a study by Silkhi et al. [18] has compared existing development tools, such as BotPress, DialogFlow, and Rasa Framework, using only technical metrics and focusing on platform capabilities. One of the few studies to explore the usability of chatbot development tools is presented by Nikou and Chang [14], which examined user perceptions of a block-based chatbot creation environment for educators through post-use surveys, revealing that while users recognized the educational value of such tools, they required substantial ongoing support due to interface complexity and development challenges.

The lack of investigation related to the user experience in chatbot development platforms represents a significant research gap, particularly given that the usability of development tools directly impacts both the quality of resulting chatbots and the institutional adoption of such tools. The present study addresses this gap by providing a longitudinal examination of user experience evolution with an educational chatbot creation platform.

3 Methodology

To address the identified research gap, we conducted a longitudinal diary study examining how users' perceptions and interactions evolve over extended use of an educational chatbot creation platform. This section describes the platform used for our study, the research design, participant recruitment, and data collection procedures.

3.1 Study Platform

This study was conducted in the context of the BUas AI Assistant Platform, a web-based chatbot creation system developed through an internal research project at Breda University of Applied Sciences. The platform was designed to address key challenges in educational chatbot development by enabling university community members (lecturers, staff, and students) to create and experiment with conversational agents without requiring extensive technical expertise. The development was motivated by institutional needs for data governance and pedagogical alignment, utilizing locally hosted LLMs to avoid dependency on external commercial providers while maintaining control over sensitive data.

The platform's interface design draws inspiration from contemporary conversational AI systems such as OpenAI's ChatGPT, Anthropic's Claude, and Google's Gemini, particularly in its chat interface components to ensure user familiarity. However, unlike these general-purpose conversational tools, the platform focuses specifically on the creation and customization of specialized assistants rather than direct interaction with pre-configured models. The platform implements a structured approach to chatbot development that decomposes the complex process of prompt engineering into five guided components: identity and purpose definition, expertise specification, capability configuration, operational

restrictions, and additional behavioral instructions. This decomposition strategy was designed to make LLM prompt engineering accessible to users with varying technical backgrounds while maintaining standard chatbot design practices [3].

A progressive disclosure pattern is employed in the main interface of the platform, presenting users with a familiar dashboard for browsing existing chatbots organized by visibility categories (Fig. 1 (a)), followed by a guided form-based workflow that allow users to create their chatbots (Fig. 1 (b)). Users can create chatbots with knowledge bases by uploading files in multiple formats, and select the LLMs to use from locally hosted open-source models (LLaMA, Qwen, DeepSeek) or configure API access for proprietary models (OpenAI's GPT models). The platform provides three deployment visibility levels (private, unlisted, and public), enabling iterative testing and collaboration workflows.

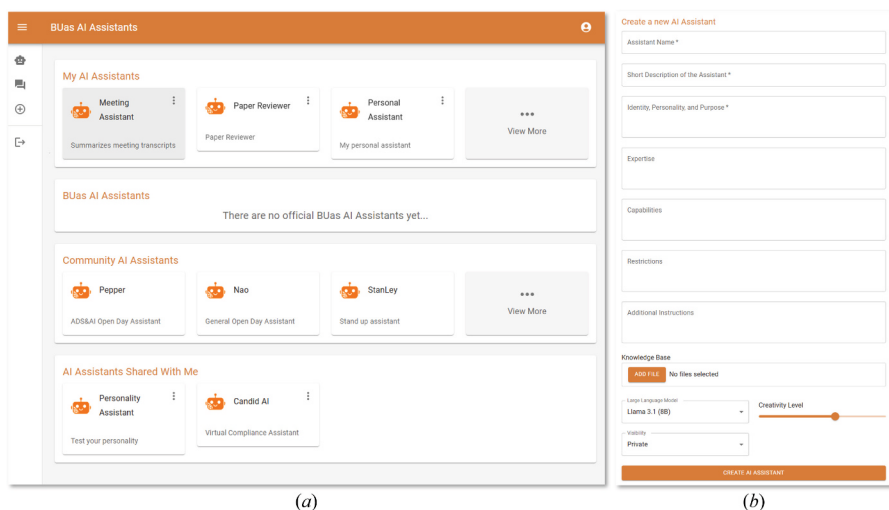


Fig. 1. User Interface of the BUas AI Assistant Platform.

3.2 Study Design and Task Context

The study was conducted within a two-week academic challenge that is part of the Applied Data Science and Artificial Intelligence program curriculum at Breda University of Applied Sciences. These challenges are designed as autonomous learning experiences where students from years 1 and 2 are presented with real-world problems proposed by external companies and are expected to develop solutions independently with minimal guidance or instructor assistance. For this particular challenge, participants were tasked with creating a Virtual Compliance Assistant chatbot for a company focused on workplace integrity and compliance training.

Students received detailed documentation about anti-bribery and corruption policies, specifically gift policies, and were required to develop a chatbot capable of accurately answering employee queries about compliance guidelines. The challenge required participants to engage with multiple complex aspects of chatbot development: analyzing and comprehending detailed policy documentation, structuring knowledge bases for effective information retrieval, crafting appropriate system prompts using the platform's structured components, and iteratively testing and refining chatbot responses to ensure accuracy and policy alignment. At the conclusion of the two-week period, students were required to present their solutions.

This task design provided an goal-oriented use case that required continuous usage of the platform throughout the study period. The independent nature of the challenge ensured that participants needed to explore and learn the platform's capabilities autonomously, providing ideal conditions for observing natural user experience evolution and adaptation processes.

We employed a diary-based research methodology to capture user experience evolution during this extended engagement with the system. Diary studies are a qualitative user research method that enables the collection of insights about user behaviors, activities, and experiences over time [1]. This longitudinal approach was selected to capture how user perceptions and interactions with the platform evolved during extended use, providing an in-depth understanding of usability challenges and user adaptation processes that would not be visible in single-session evaluations.

3.3 Participants

Participants were recruited from first and second-year students enrolled in the Applied Data Science and Artificial Intelligence program of Breda University of Applied Sciences who demonstrated interest on participating in the two-week academic challenge described above. All students participating in the challenge were invited to voluntarily join the user study.

The study population consisted of 10 students (7 males, 3 females, $M_{age} = 19.8$ years old, range = [18,21]) with varying levels of technical expertise, ranging from those with limited programming experience to more advanced students familiar with machine learning and natural language processing concepts. This diversity was particularly relevant to our research objectives, as the platform was specifically designed to accommodate users across different technical proficiency levels. All participants provided informed consent prior to study commencement, acknowledging their voluntary participation and the right to withdraw at any time without consequences affecting their academic standing.

The study received full institutional approval from Breda University of Applied Sciences. All participants provided informed consent acknowledging voluntary participation and withdrawal rights. Data confidentiality was ensured through encrypted storage on secure servers accessible only to authorized research team members, with all personally identifiable information

removed from analysis datasets. Participants were informed that their individual responses would remain confidential and would not influence their academic assessment in any way.

3.4 Data Collection Procedure

The study followed a structured three-phase data collection approach spanning two weeks. Participants completed diary entries via Microsoft Forms at strategically selected intervals to capture different stages of their platform experience: immediately after their first day of interaction (Diary 1), after one week of sustained use (Diary 2), and after completing the full two-week challenge (Diary 3). This timing was designed to capture initial impressions, adaptation processes during extended use, and final reflections on the complete experience.

Each diary session required approximately 30 min to complete and combined multiple data collection methods. The diary entries incorporated three primary data types: qualitative open-ended questions exploring user experiences, platform impressions, perceived strengths and weaknesses, and suggestions for improvement; quantitative usability measures adapted from the Unified Theory of Acceptance and Use of Technology (UTAUT) framework [19] focusing on perceived ease of use, interaction clarity, and skill development progression; and demographic information (age, gender, prior chatbot development experience). Participants were instructed to provide detailed, reflective responses rather than brief answers.

Participants received reminder notifications for each diary phase and were provided with clear instructions emphasizing the importance of honest and detailed feedback for improving the platform's effectiveness.

4 Results and Discussion

4.1 Participants' Backgrounds

The analysis of participants' backgrounds revealed diverse prior experience with chatbot creation. Half of the participants ($n = 5$) reported having no previous experience with chatbot development, while the remainder had experimented with various platforms including open-source chatbot builders, web-based chatbot tools such as Character AI, and development frameworks such as Ollama.

4.2 Initial Platform Encounters and Learning Trajectory

User satisfaction with the platform showed a consistent upward trajectory over the study period. Participants rated their overall satisfaction on a 10-point scale at each diary phase, with mean ratings increasing from 7.2 ($SD = 0.75$) after the first day of use to 7.7 ($SD = 0.90$) after one week, and reaching 8.2 ($SD = 0.87$) by the end of the two-week period. While this upward trend suggests improving user satisfaction with extended platform exposure, a repeated measures ANOVA revealed the change was not statistically significant ($F(2, 27) = 3.17, p = 0.058$).

possibly due to the small sample size. This pattern suggests that initial user reactions to the platform, while generally positive (mean > 7), continued to improve as participants gained familiarity with the chatbot creation process. The increasing satisfaction aligns with expectations for complex tools where user appreciation typically grows with competency development, as users feel more confident in their abilities and develop more sophisticated mental models of system functionality [13].

Participants reported that during their first day of interaction, they engaged with core platform functionalities including creating basic assistants, uploading knowledge bases, and testing parameter changes. Initial user impressions revealed a mixed but generally positive reception of the platform's design. Most participants found the interface accessible and well-designed, with comments highlighting its simplicity: "The interface so far is pretty basic and quite straightforward" and "Really user friendly, clean and easy to use". Several users specifically noted the intuitive navigation, with one participant observing an "Intuitive layout that made navigation simple for me". However, some participants experienced initial uncertainty when encountering the platform's structured approach to chatbot creation. One user noted feeling "a little bit overwhelmed or intimidated by the provided boxes and their pop-up explanations", while another described the learning process as requiring exploration: "It was fun to figure things out, but it wasn't easy, some things I had to understand through trial and error".

The usage patterns evolved significantly throughout the study period, reflecting growing user engagement. During the second week, most participants used the platform 3–4 times, with usage ranging from a single session to daily interaction. By the final week, engagement had intensified considerably, with half of the participants using the platform 2–3 times per day and the other half engaging daily. This progression from occasional to frequent use suggests that participants found increasing value in the platform as they developed familiarity with its capabilities.

At the end of the study, participants reported clear skill development in both chatbot creation concepts and platform-specific competencies. Users described enhanced understanding of fundamental chatbot principles, with one participant noting: "I enhanced my understanding of aspects influencing chatbot's responses", while another reflected on more general skills: "I became more skilled in overall prompt engineering". Platform-specific skill development was also evident, as participants gained confidence in navigating the interface and utilizing its features. One user stated: "I believe I have become more skilled in using the platform", though others showed more cautious self-assessment: "I feel like I have gained a lot of insight of how the platform works but I have to work a bit more on it in order to say if I have become skilled in using the platform".

4.3 Usability Perceptions and Platform Interaction

Participants' perceptions of effort required to use the platform were measured using three UTAUT Effort Expectancy items on a 7-point scale (Cronbach's

$\alpha = 0.80$). Results showed an improvement in perceived ease of use between the first diary ($M = 5.16, SD = 0.99$) and the second diary after one week of use ($M = 5.70, SD = 0.98$), with scores remaining stable through the third diary entry ($M = 5.66, SD = 0.66$). Although a repeated measures ANOVA showed these changes were not statistically significant ($F(2, 27) = 1.11, p = 0.344$), they suggest that participants experienced the greatest reduction in perceived effort during their first week of sustained platform use, with effort expectations stabilizing thereafter. The relatively high initial scores (above the scale midpoint of 4) indicate that participants found the platform reasonably accessible from their first encounters, aligning with the design goal of supporting users across varying technical proficiency levels.

Perceived ease of use was assessed through four UTAUT items on a 7-point scale, measuring clarity of platform use, goal achievement ease, interaction understandability, and overall ease of use (Cronbach's $\alpha = 0.72$). Results demonstrated a consistent upward trajectory across the study period, with mean scores increasing from 5.55 ($SD = 0.75$) in Diary 1 to 5.90 ($SD = 0.59$) in Diary 2, and reaching 6.02 ($SD = 0.70$) by Diary 3. However, a repeated measures ANOVA revealed these improvements were not statistically significant ($F(2, 27) = 1.29, p = 0.292$). This pattern aligns with the previously reported satisfaction and effort expectancy trends. The initial mean score of 5.55 indicates that participants found the platform moderately easy to use from their first encounters, suggesting successful initial usability design. The progressive increase to above 6.0 by the final diary represents movement toward the higher end of the ease-of-use spectrum, indicating that sustained interaction with the platform led to improved user confidence and perceived usability.

These usability perceptions were reflected in participants' evolving interactions with the platform. While initial actions focused on fundamental exploration, later interactions during the second week became more sophisticated and targeted, shifting toward adjusting prompts and refining parameter configurations to create more optimized chatbots.

Participants recognized their developing competency in platform use, particularly in prompt engineering skills. Users described becoming more strategic in their approach: "Yes, by writing better prompts" and "Yes, I feel like my understanding became more clear and I was able to manipulate more things with ease". One participant provided detailed insight into their learning progression: "Yes, definitely. I learned which boxes are the most important for defining the chatbot, and how to put emphasis on the most important instructions for the chatbot to respect them. I also learned the difference between the private and public chatbots, different creativity levels and tested different types of language models".

By the study's conclusion, participants reflected positively on their overall platform experience, consistently emphasizing its user-friendly design. Comments included: "The platform was super easy and straight forward and the challenge didn't take me a lot of time" and "I liked using this platform. It was a nice experience mostly because of how user friendly the it is". Notably, some

participants acknowledged that their initial negative impressions had been influenced by inexperience: “I feel like my previous feelings about the platform were skewed towards the negatives because of my lack of knowledge about using this site”.

Regarding goal achievement, eight of ten participants felt they had accomplished their intended objectives with the platform. The two participants who expressed reservations cited specific technical limitations: one noted “Not necessary, I lacked the deeper understanding to do what I envisioned” while another identified model accuracy concerns: “I did not achieve everything I wanted to with this platform because hallucinations still occur sometimes. I believe an improved model would help create a more accurate bot”.

4.4 Platform Strengths and User Value

Participants consistently highlighted specific platform attributes that contributed to their positive experience, with these perceived strengths evolving and deepening over the course of their engagement. Initial impressions focused primarily on fundamental usability characteristics, with users emphasizing the platform’s straightforward design: “It is very clear, straightforward and easy to use” and “It’s intuitive, I didn’t have to spend too much time figuring out how to do things”.

Early appreciation also centered on resource accessibility, particularly the availability of multiple language models without additional cost barriers. Participants valued this institutional support, with one noting: “We are being provided by a free language model by the university, which is very advantageous for a first-time user”. The model selection flexibility was also appreciated: “I liked the fact that there are multiple AI models to choose from.”

By the study’s conclusion, participants had developed a more sophisticated appreciation for the platform’s advanced capabilities beyond basic usability. Users began recognizing the creative potential enabled by the structured approach: “I would say infinite possibilities for creativity in designing chatbot” and “The customization capabilities are one of the best aspects of the platform”.

Interestingly, while the fundamental ease of use remained consistently valued throughout the study period, participants’ final reflections demonstrated deeper understanding of the platform’s design philosophy. Comments such as “This platform’s most valuable aspect was its ease of understanding and control” and “It was really easy to follow the steps needed to create the chatbot” suggest that users had come to appreciate not just surface-level usability, but the underlying structure and design choices that made complex chatbot creation manageable.

This evolution from basic usability appreciation to recognition of sophisticated customization capabilities illustrates how sustained platform engagement revealed layers of value that were not immediately apparent during initial encounters. The perceived value was sufficient that most participants expressed interest in future platform use, particularly for academic purposes, with one noting: “I think it’s very nice for practicing how to build a chatbot. Additionally, it’s interesting to try various possible ways to build a chatbot with it”.

4.5 Areas for Improvement

Despite the overall positive user experience, participants consistently identified specific areas requiring improvement throughout the study period. These challenges remained largely stable across diary entries, suggesting persistent usability barriers.

The most frequently cited improvement need centered on providing clearer guidance and instructional support for new users. Participants requested more fundamental instructions to understand the platform's prompt engineering approach, suggesting: "It would benefit from placeholders in prompt windows to show what type of prompt should be written" and "Mainly the explanations could be improved and providing more examples would be nice to see". These requests indicate that while users appreciated the platform's structured approach, they desired more explicit guidance on how to effectively create effective chatbots.

Participants also identified specific functionality enhancements that could improve the creation workflow. Practical suggestions included interface improvements such as making "the textboxes enlarge with the size of the text for easy editing" and system performance enhancements such as "making the generation process faster".

By the second diary, participants began expressing interest in more sophisticated platform capabilities. One user noted that "the platform could have also some more advanced options which would make it even more interesting to use and learn with" while other provided more general feedback that "the UX/UI could be improved" without specifying particular aspects.

Notably, the improvement suggestions identified in the final diary largely repeated those from earlier entries, indicating that while users' appreciation for the platform grew over time, their perception of its limitations remained consistent. This suggests that these usability challenges represent fundamental design opportunities rather than issues that resolve through extended use.

5 Concluding Remarks

In this study, we provided a longitudinal examination of user experience evolution with an educational chatbot creation platform. Through a two-week diary-based study with students using the BUas AI Assistant Platform, we demonstrated that user experience with complex AI creation tools follows predictable temporal dynamics that single-session evaluations cannot capture.

Our findings reveal consistent improvement trajectories across multiple usability dimensions – satisfaction, effort expectancy, and perceived ease of use – alongside qualitative evidence of skill development and deepening platform appreciation. Users progressed through a two-phase learning curve: initial exploration marked by mixed impressions and uncertainty, followed by competency consolidation characterized by increased engagement and sophisticated feature recognition. Crucially, certain usability barriers persisted throughout the study period, indicating fundamental design issues that are not resolved through familiarity alone. For the factors influencing user experience trajectories, we identified

that structured decomposition of complex prompt engineering into guided components, combined with authentic task contexts requiring sustained engagement, were the primary determinants of successful user adaptation and skill development.

The study revealed several key insights. First, longitudinal evaluation is essential for complex AI systems, as initial moderate usability ratings improved considerably over extended use periods. Second, authentic task contexts requiring sustained engagement are crucial for revealing the full user experience trajectory, as superficial feature testing cannot capture how users actually adapt to complex AI tools. Third, successful AI creation platforms must explicitly design for two-phase learning trajectories, providing progressive disclosure that prevents initial overwhelm while enabling capability growth. Finally, decomposing complex AI concepts into structured, guided frameworks can make sophisticated prompt engineering accessible to users with varying technical backgrounds without sacrificing functionality.

Our study's exploratory nature with 10 participants provided rich qualitative insights but limits generalizability. Future research should examine larger, more diverse user populations across different educational contexts and task types. Additionally, investigating longer-term usage patterns beyond two weeks could reveal further evolution phases in user experience trajectories. Comparative studies evaluating different design approaches to prompt engineering guidance would help validate our design recommendations.

The findings of this study establish a foundation for understanding temporal dynamics in chatbot creation tools, providing actionable frameworks for designing educational AI platforms that support meaningful user learning trajectories. As educational institutions increasingly adopt chatbot creation platforms, this research offers evidence-based guidance for developing systems that balance initial accessibility with long-term capability growth, ultimately enabling more effective integration of AI technologies in educational contexts.

The findings of this study establish a foundation for understanding temporal dynamics in chatbot creation tools, providing actionable frameworks for designing educational AI platforms that support meaningful user learning trajectories. As generative AI continues to expand into diverse applications, from healthcare diagnostics and medical education [5,17] to narrative generation and interactive storytelling [10–12], understanding how users adapt to AI creation tools becomes increasingly critical for effective technology adoption. As educational institutions increasingly adopt chatbot creation platforms, this research offers evidence-based guidance for developing systems that balance initial accessibility with long-term capability growth, ultimately enabling more effective integration of AI technologies in educational contexts.

References

1. Bolger, N., Davis, A., Rafaeli, E.: Diary methods: capturing life as it is lived. *Annu. Rev. Psychol.* **54**, 579–616 (2002)
2. Borsci, S., et al.: The Chatbot usability scale: the design and pilot of a usability scale for interaction with ai-based conversational agents. *Pers. Ubiquit. Comput.* **26**(1), 95–119 (2021). <https://doi.org/10.1007/s00779-021-01582-9>
3. Chaves, A.P., Gerosa, M.A.: How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *Int. J. Hum.-Comput. Int.* **37**(8), 729–758 (2021). <https://doi.org/10.1080/10447318.2020.1841438>
4. Chen, L., Chen, P., Lin, Z.: Artificial intelligence in education: a review. *IEEE Access* **8**, 75264–75278 (2020). <https://doi.org/10.1109/ACCESS.2020.2988510>
5. Chen, Y., Esmaeilzadeh, P.: Generative ai in medical practice: in-depth exploration of privacy and security challenges. *J. Med. Internet Res.* **26**, e53008 (2024). <https://doi.org/10.2196/53008>
6. Dagkoulis, I., Moussiades, L.: A comparative evaluation of chatbot development platforms. In: *Proceedings of the 26th Pan-Hellenic Conference on Informatics*, pp. 322–328. PCI '22, Association for Computing Machinery, New York, NY, USA (2023). <https://doi.org/10.1145/3575879.3576012>
7. Hennick, C.: How three universities developed their chatbots. *EdTech Magazine (Higher Education)* (2025). <https://edtechmagazine.com/higher/article/2025/05/how-three-universities-developed-their-chatbots>. Accessed 08 Sept 2025
8. Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., Mctear, M.: Usability testing of a healthcare chatbot: can we use conventional methods to assess conversational user interfaces? In: *Proceedings of the 31st European Conference on Cognitive Ergonomics*, pp. 207–214. ECCE '19, Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3335082.3335094>
9. 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). <https://doi.org/10.1186/s41239-023-00426-1>
10. de Lima, E.S., Feijó, B., Casanova, M.A., Furtado, A.L.: Chatgeppetto - an AI-powered storyteller. In: *Proceedings of the 22nd Brazilian Symposium on Games and Digital Entertainment*, pp. 28–37. ACM (2024). <https://doi.org/10.1145/3631085.3631302>
11. de Lima, E.S., Neggers, M.M.E., Casanova, M.A., Feijó, B., Furtado, A.L.: A pattern-oriented AI-powered approach to story composition. In: Figueroa, P., Di Iorio, A., Guzman del Rio, D., Gonzalez Clua, E.W., Cuevas Rodriguez, L. (eds.) *Entertainment Computing – ICEC 2024*, pp. 1–16. Springer Cham (2024). https://doi.org/10.1007/978-3-031-74353-5_10
12. de Lima, E.S., Neggers, M.M., Feijó, B., Casanova, M.A., Furtado, A.L.: An AI-powered approach to the semiotic reconstruction of narratives. *Entertain. Comput.* **52**, 100810 (2025). <https://doi.org/10.1016/j.entcom.2024.100810>
13. Nielsen, J.: *Usability Engineering*. Morgan Kaufmann, San Francisco (1993)
14. Nikou, S.A., Chang, M.: Learning by building chatbot: a system usability study and teachers' views about the educational uses of chatbots. In: *Augmented Intelligence and Intelligent Tutoring Systems: 19th International Conference, ITS 2023, Corfu, Greece, June 2–5, 2023, Proceedings*, pp. 342–351. Springer-Verlag, Berlin, Heidelberg (2023). https://doi.org/10.1007/978-3-031-32883-1_31

15. Peras, D.: Chatbot user experience: Design and evaluation. In: Coman, A., Vasilache, S., Fui-Hoon Nah, F., Siau, K.L., Wei, J., Margetis, G. (eds.) *HCI International 2024 – Late Breaking Papers*, pp. 77–93. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-76806-4_6
16. Rapp, A., Curti, L., Boldi, A.: The human side of human-chatbot interaction: a systematic literature review of ten years of research on text-based chatbots. *Int. J. Hum Comput Stud.* **151**, 102630 (2021). <https://doi.org/10.1016/j.ijhcs.2021.102630>
17. Reddy, S., et al.: Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implement. Sci.* **19**(1), 27 (2024). <https://doi.org/10.1186/s13012-024-01357-9>
18. Silkhi, H., Bakkas, B., Housni, K.: Comparative analysis of rule-based chatbot development tools for education orientation: a rad approach. In: *Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security*. NISS '24, Association for Computing Machinery, New York, NY, USA (2024). <https://doi.org/10.1145/3659677.3659825>
19. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. *MIS Q.* **27**(3), 425–478 (2003). <https://doi.org/10.2307/30036540>
20. Yigci, D., Eryilmaz, M., Yetisen, A.K., Tasoglu, S., Ozcan, A.: Large language model-based chatbots in higher education. *Adv. Intell. Syst.* **7**(3), 2400429 (2025). <https://doi.org/10.1002/aisy.202400429>