Integrating Gamification, Recommender System, and Chatbot in Cybersecurity Training: A User-Centered Approach for Enhanced Engagement and Motivation

Ling Leung[1], Benoît Duhoux[1], Axel Legay[2], and Suzanne Kieffer[1]

[1] Université catholique de Louvain, Louvain-la-Neuve, Belgium
{ling.leung,benoit.duhoux,suzanne.kieffer}@uclouvain.be
[2] Nexova, Libin, Belgium
a.legay@nexovagroup.eu

Abstract. Cyber attacks are increasing, causing financial losses, data breaches, and reputational damage for individuals and organizations. Since human error remains the leading cause, raising awareness of cyber threats is essential. Cyber range scenarios aim to mitigate these risks but often fail to reflect real-world challenges and user needs. This underscores the importance of integrating human-computer interaction (HCI) and user experience (UX) principles to develop more effective cybersecurity training. While gamification, recommender system, chatbot enhance the user experience in cybersecurity training, no existing system seamlessly combines these components. This paper explores the user requirements for such a system using a prospective user-centered approach. Ten potential users participate in tests involving task-based exercises, UEQ+ assessment, ad-hoc questionnaire, and interviews. The results suggest that integrating these components into a unified system can significantly enhance user engagement in cybersecurity training.

Keywords: Gamification · Chatbot · Recommender system · Cybersecurity training · Cyber range · User engagement · User experience · Human-computer interaction

1 Introduction

Cybersecurity is a multidisciplinary field with significant impacts on economic, environmental, and social systems worldwide. Recent statistics indicate that 2,200 cyberattacks occur daily, averaging one every 39 seconds [14].

As organizations increasingly rely on computerized systems to manage their operations, they become more vulnerable to cyberattacks. Human error remains the leading cause of these breaches [46], underscoring the urgent need for effective mitigation strategies. With cyber threats evolving continuously, cybersecurity has become a critical priority for all digital technology users. However, a major obstacle in cybersecurity education is the lack of realistic training environments.

Unlike fields such as aerospace or medicine, cybersecurity often lacks simulation platforms necessary for effective skill development.

Cyber ranges address this gap by providing interactive platforms that replicate “an organization’s ICT, OT, mobile and physical systems, applications, and infrastructures, including the simulation of attacks, users and their activities, and any other Internet, public or third-party services” in a secure and controlled setting [8]. By offering a safe and legal space for hands-on learning, cyber ranges enable organizations to test their security measures and strengthen their defenses against cyber threats. Combining physical and virtual components, these platforms are essential for education, workforce development, and certification.

However, cyber ranges face several challenges, including maintaining motivation for ongoing education, particularly among non-specialists, as well as the lack of adaptive learning difficulty, real-time scenario support, personalized guidance, and scenarios tailored to individual needs [46].

Gamification, chatbot, and recommender system each address these issues. Gamification, through elements like scoring, badges, and progress bars, enhances engagement, motivation, and learning performance [42], fostering achievement, competition [24], and improving training outcomes [16]. Chatbot improves learning outcomes and knowledge retention by allowing users to ask questions, receive real-time hints, and benefit from adaptive feedback and personalized training, keeping them engaged [22]. Recommender system provides personalized content and suggestions, adapting to user preferences, competencies and performance to boost engagement, satisfaction, and training relevance [49]. Combining these three components addresses their limitations and boosts user motivation and engagement. To our knowledge, no cybersecurity system integrates them in a unified manner. The primary challenge lies in designing a system that effectively orchestrates these components, ensuring seamless interaction without compromising user experience.

The main contribution of this paper is twofold: it specifies user requirements and tests an interactive prototype of the proposed solution. Using a user-centered approach, it identifies specific requirements to create an engaging and effective learning experience and it tests a first version of the prototype to fix usability, user experience issues and assess the user requirements. Rather than applying inferential statistics, we employed a qualitative test-and-improve approach [18], focusing on iterative refinement through user feedback. This method prioritizes in-depth insights from a representative sample over generalized conclusions, ensuring the platform’s design and functionalities align with actual user needs, rather than relying solely on quantitative metrics or causal modeling. By defining key requirements and detailing the iterative development process, this study lays the groundwork for a cybersecurity training solution that optimizes engagement, usability, and learning effectiveness.

2 Background and related work

Gamification. Sustaining trainee motivation and focus remains a significant challenge in e-learning research, as trainees often lose interest quickly [11]. Researchers have turned to gamification, a proven solution for maintaining engagement during learning activities [41]. Gamification enhances traditional training methods by integrating game design elements into non-game contexts [4, 6, 13, 15]: points, leaderboards, badges, and progress tracking [2]. These mechanisms are particularly effective in cybersecurity education, where technical concepts can sometimes feel abstract. By turning complex challenges into engaging activities, gamification fosters both extrinsic and intrinsic motivation [26], encourages perseverance, and helps trainees stay focused. The effectiveness of gamification lies in its ability to simplify learning from both cognitive and emotional perspectives, automatic feedback, and personalized learning experience [36]. Above all, it leads to greater user engagement and participation in online courses [3]. However, a "one-size-fits-all" approach is unlikely to meet the diverse needs of all users due to individual differences [37, 47]. Consequently, gamification often fails to provide tailored guidance or address the specific needs of individual trainees.

Recommender system. Recommender system personalizes cybersecurity training by analyzing user interactions, needs, interests [25], preferences, and performance to suggest relevant learning scenarios, such as password management or phishing detection [33]. Their adaptability ensures that trainees remain appropriately challenged, dynamically adjusting training based on user progress to prevent both disengagement and cognitive overload [33]. Additionally, they address common challenges such as cold start, data sparsity, scalability, and the balance between diversity and accuracy [25]. However, while recommender system optimizes content selection, it does not guarantee engagement or long-term knowledge retention. It lacks real-time guidance, which may leave trainees struggling to understand recommendations or complete tasks effectively. Without chatbot support, users may find it difficult to grasp why certain recommendations are made or how to successfully complete suggested activities. Furthermore, recommender system does not inherently motivate trainees.

Chatbot. Chatbot technology, powered by advances in large language models like GPT-4 and Gemini, is transforming education by enabling sophisticated linguistic interactions [48]. Multiple studies demonstrate their effectiveness across various educational contexts [5, 7, 12, 21, 28–32, 35, 44], highlighting their potential to improve learning outcomes and enhance trainee satisfaction [32]. In cybersecurity training, where threats evolve rapidly, chatbots serve as interactive mentors by providing instant feedback [1], answering questions [43], and guiding users through complex tasks. Their ability to personalize interactions, by adapting responses to users' needs, actions, and emotions [9, 10, 19, 38], engages both novice and advanced trainees. Integrating mechanisms such as retrieval-augmented generation enhances their capacity to deliver domain-specific insights. For example, during a simulated phishing attack, trainees can receive targeted hints without disrupting their learning flow [46].

However, chatbot alone often falls short of maintaining engagement. Their limitations stem from insufficient dataset training, reducing adaptability and frustrating users [20, 22]. Furthermore, issues such as poor user-centered design, inadequate feedback, and potential distractions further diminish their educational effectiveness [34].

3 Proposed solution

The proposed solution integrates gamification, a recommender system, and a chatbot into a unified system, leveraging their respective strengths while mitigating their respective drawbacks, thereby increasing engagement in cybersecurity education. This section presents the user requirements and the prototype.

3.1 Requirements

As the system integrates gamification, recommender system, and chatbot, it must meet the specific requirements of each. However, these requirements are refined to emphasize how the synergy between these components is essential for maximizing engagement and mitigating their individual limitations.

Gamification is important for stimulating engagement and fostering long-term commitment, meaning that it must incorporate elements such as scores, badges, and leaderboards (R1) to promote a sense of achievement and motivate users to actively engage in their learning process. The gamified elements must also enable users to visualize their progress through various visual indicators (R2), helping them track their learning path and encouraging sustained effort.

The recommender system must suggest training scenarios tailored to users' competencies, preferences, and learning objectives (R3) while refining recommendations based on data from gamified elements (scores, badges) and chatbot interactions. Frequent assistance requests or recurring difficulties can signal the need for less complex scenarios. User preferences and evolving goals, whether explicitly set or inferred through the chatbot, further shape a personalized learning path. Additionally, the system must ensure coherent narrative progression, considering completed scenarios and guiding users through interconnected missions aligned with their objectives (R4), such as advancing toward a junior penetration tester role.

The chatbot must act as an adaptive guide, providing personalized real-time advice based on scenarios from the recommender system and the user's selected difficulty level (R5). It must also help users track progress, offering a clear view of achievements and areas for improvement (R6) while adjusting assistance based on scenario complexity and user performance metrics from gamified elements (R7), ensuring support without undermining autonomy. When users struggle, the chatbot must deliver contextualized hints—neither too obvious nor too vague (R8)—to maintain the educational value of problem-solving. Finally, it must dynamically adapt storytelling to reflect users' evolving learning objectives and progress (R9), keeping the training experience relevant and engaging.

3.2 Prototype

For the sake of brevity, we focus on and illustrate only core features of the prototype. Upon accessing the platform, users are directed to their "Dashboard", shown at the top of Figure 1. The top section displays a profile summary with personal information such as username, total score, and current company affiliation. Additionally, a skills matrix offers an overview of proficiency levels across different fields. Beyond these initial gamified elements—total score and skills matrix—further statistics are available at the bottom of the page, allowing users to track their progress by difficulty level or scenario type. More detailed statistics can be accessed via the "View Advanced Statistics" button.

In the middle of the dashboard page, users can set learning objectives or update their preferences to fine-tune the scenarios recommended by the system. These personalized recommendations appear on the "Scenarios" page, illustrated at the bottom of Figure 1, alongside recently played scenarios and additional challenges. Each scenario displays essential information, including its name, difficulty level, associated domains, overall ratings, and the user's personal score. Users can access more details via the "Details" button and launch the scenario using the "Start" button. To accommodate users who prefer a more flexible learning approach, the system offers a "Filter Scenarios" panel on the left-hand side, allowing exploration of the full scenario list. Scenarios can be filtered by date, difficulty, type, or top-rated options.

To define a learning goal, users navigate to the "Goals" page, accessible via the navigation bar or the "Your Learning Goals" panel on the dashboard. This page provides an overview of the user's selected objectives and scenario preferences, as shown on the dashboard. If no objectives have been defined, users can explore predefined learning paths in the "Your Path to Success" panel. For example, a user aiming to become a junior penetration tester can select this path, review its roadmap and difficulty level, and enroll if it matches their goals.

The system also fosters competition through the "Leaderboard" page, where users can view global and monthly rankings. Additionally, users can collect achievements in the form of badges, accessible via the "Badges" page. These badges are awarded upon reaching specific milestones, such as completing ten scenarios or attaining a total score of 500.

In the bottom right corner of each page, a chatbot provides interactive assistance, adapting its messages based on the page the user is viewing. For instance, on the "Leaderboard" page, the chatbot displays the user's current ranking along with encouraging messages to boost motivation, as illustrated in Figure 2. When a user enrolls in a learning path, the chatbot suggests adjusting scenario preferences to align with the chosen path, as depicted in Figure 3. Finally, when users encounter difficulties in a scenario, they can request hints or additional explanations to help them progress.

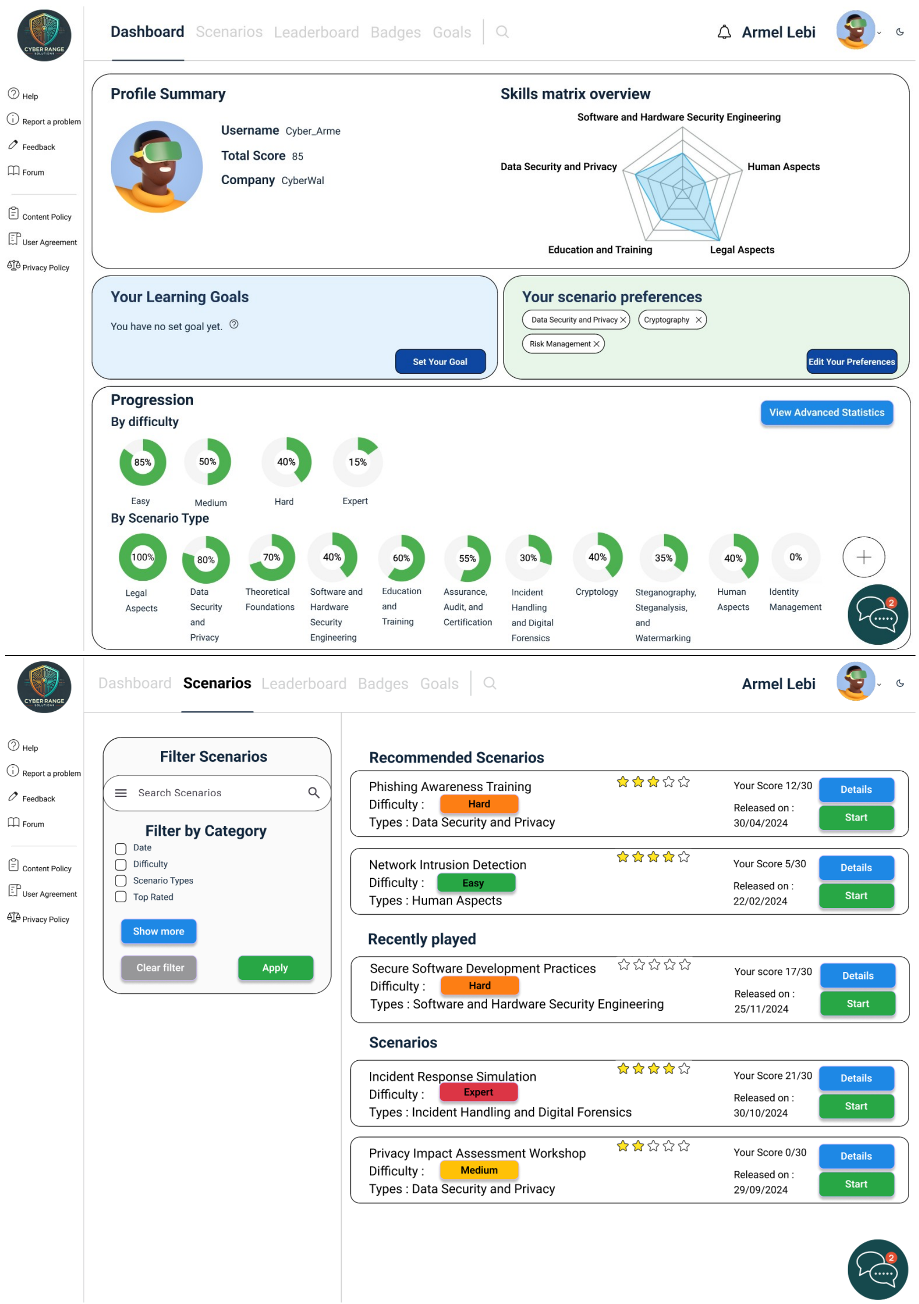

Fig. 1. The top wireframe shows the "Dashboard" page and the bottom the "Scenarios" page.

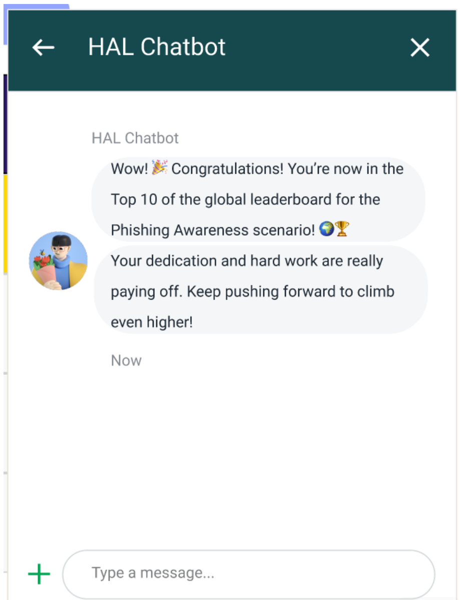

Fig. 2. Screenshot showing chatbot interaction when users access the "Leaderboard" page.

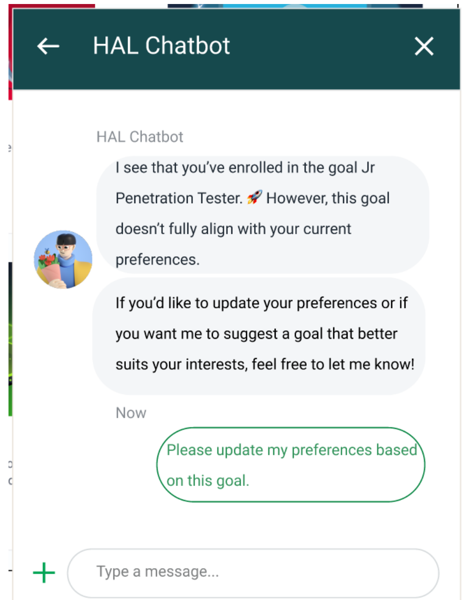

Fig. 3. Screenshot showing chatbot interaction when users enrolled for the "Junior Penetration Tester" training objective.

4 User testing methodology

4.1 Objective

The objective of the user testing is twofold: (1) to demonstrate the feasibility of the proposed integration and (2) to evaluate the system requirements through prototype testing. This paper focuses on the first iteration of a test-and-refine methodology, which is commonly used in user-centered design [18]. The process analyzes the proposed system based on how users perceive and experience it after interacting with the prototype. The feedback collected during testing informs updates to the prototype and adjustments to the system requirements. This iterative cycle continues in future iterations until the user requirements are fully satisfied. The user testing process in this first iteration includes several key components. It begins with seven usage scenarios, specifically designed to simulate real-world use cases. Participants then complete a standardized questionnaire evaluating the user experience. Additional Likert-scale questions gather more specific insights. Finally, a debriefing session provides an opportunity for participants to share qualitative feedback and discuss their overall impressions.

4.2 Participants

We recruit ten participants (aged 25 to 35, including two women) from the Institute of Information and Communication Technologies, Electronics and Applied Mathematics laboratory at UCLouvain. All participants are researchers in cybersecurity, either doctoral candidates or post-doctoral fellows. This recruitment follows a convenience sampling strategy, which allows us to easily access individuals from the target population. To ensure a reasonable level of internal validity, none of the participants have prior exposure to the prototype before the testing sessions.

4.3 Consent form

Participants sign a standard consent form that complies with the General Data Protection Regulation requirements and aligns with the ethical principles outlined in [17]. The form explains the research objectives, potential risks, mitigation measures, and how the data is used and protected. It informs participants of their rights to access, correct, or delete their data and assures them they can withdraw at any time. The study collects only the data strictly necessary and stores it securely. Only authorized researchers access the data, and it is retained only for the time needed to achieve the research objectives.

4.4 Scenarios of use

The user tests focus on seven scenarios that evaluate specific features of the prototype. These scenarios are primarily designed to validate the system's functionality (Objective 2), while questionnaires and a debriefing session assess the feasibility of the proposed integration (Objective 1). Table 1 presents the seven scenarios and their mapping to the user requirements. The scenarios cover a range of interactions, including modifying preferences, retrieving progress statistics, interacting with the chatbot, starting scenarios in expert mode, and exploring gamification elements like leaderboards, objectives, and badges.

While the current set of scenarios addresses most user requirements, three remain untested in this iteration. First, the system's ability to dynamically adjust recommendations based on interaction data, such as the frequency of help requests and encountered challenges (R3), is not explicitly evaluated. Although Scenarios 1 and 6 partially cover this requirement, they do not fully capture the recommendation system's adaptive capabilities in response to real-time user behavior. Second, the chatbot's ability to provide contextualized hints that are neither too obvious nor too obscure (R8) is also untested. This requirement demands a more complex experimental setup, which will be addressed in future iterations. Finally, the chatbot's capacity to dynamically modify the narrative structure to align with evolving user objectives and requirements (R9) remains untested. Validating this advanced feature would necessitate complex, multi-step scenarios involving iterative interactions and evolving system states.

Table 1. Mapping between scenarios of use and user requirements

Scenario	Mapped User Requirements
S1. Modify scenario preferences by adding "Legal Aspects".	R3. The recommendation system must propose scenarios adapted to the user's preferences, capabilities, and objectives.
S2. Retrieve progress statistics for the "Data Security and Privacy" scenario.	R2. Gamification must enable users to visualize their evolution and motivate them to follow the recommended path. R6. The chatbot must allow the user to track their progress in real-time.
S3. Ask chatbot about statistics of hints.	R5. The chatbot must provide personalized guidance in real-time.
S4. Start "Phishing Awareness Training" scenario in expert mode.	R4. The recommendation system must propose scenarios based on a coherent narrative progression, including user objectives and prior completions.
S5. Check both monthly and global leaderboards of users.	R1. Gamification (points, badges, leaderboards) must reinforce user engagement and motivation.
S6. Add objective "Jr Penetration Tester" and update preferences accordingly.	R3. The recommendation system must propose scenarios adapted to user objectives.
S7. Retrieve statistics for the "Cryptology" badge.	R2. Gamification must enable users to visualize their evolution and motivate them to follow the recommended path. R7. The chatbot must adjust its assistance level based on gamified performance (e.g., scores, badges).

This iteration prioritizes testing core functionalities and gathering foundational insights. The validation of advanced features, such as dynamic narrative modification and adaptive recommendations, is deferred to future iterations to ensure a focused and methodical evaluation process.

4.5 Questionnaires

We use the User Experience Questionnaire Plus (UEQ+), a standardized tool for measuring UX with up to 26 scales [40]. For this study, we select eight scales: attractiveness, efficiency, perspicuity, dependability, stimulation, novelty, quality of content, and trustworthiness of content. The first six correspond to the original UEQ scales [23], a validated and widely recognized tool that ensures reliable and comparable results across studies. The last two are specifically included to evaluate the quality and trustworthiness of content. Each scale uses four items rated on a 1 to 7 scale, along with a relative importance rating. The mean UX score for each scale and an overall UX score for the prototype are calculated from participant ratings. The mean relative importance is computed as the average importance rating across participants.

We also use an ad-hoc questionnaire with 5-point Likert-scale questions to gather detailed insights into participants' perceptions of the gamification, the recommender system, the chatbot, and their integration. First, we assess the engagement and usefulness of gamified elements, their ability to support a progressive learning path, the motivational impact of progress visualization, and the effectiveness of points and badges in increasing engagement. Second, we evaluate the relevance of the scenarios generated by the recommender system and their alignment with users' progress and difficulties. Third, we examine the quality of personalized advice provided by the chatbot, its ability to help users track progress in real time, and its adaptability to user preferences and performance. Finally, we analyze the overall integration of these three components, the consistency between recommended scenarios and users' needs and objectives, the system's effectiveness in meeting learning and progress expectations, and the presence of any inconsistencies between components.

4.6 Interviews

Following the completion of the questionnaires, we ask participants to explain their scores to better understand their opinions. Additionally, we ask three questions to gain further insights: whether they find any aspects of the system particularly innovative or distinctive compared to other solutions they have used (and if not, why), what they consider to be the main benefits of the system for a user like them, and whether they find the narrative progression coherent and motivating.

4.7 Apparatus and procedure

User testing sessions take place online via Teams, using a screen recording tool to capture participants' interactions with the prototype. To ensure full visibility without vertical scrolling, we require participants to use a dual-screen setup, with at least one 15-inch screen for the prototype and the other for scenario instructions. We also ask them to think aloud, explaining their actions, verbalizing their thoughts and feelings as they perform the scenarios. The recordings allow us to collect performance data, including scenario completion time, completion rates, and help requests. Since Teams automatically generates transcripts, we can efficiently analyze participants' verbal feedback without the need for manual transcription, saving time and effort. To encourage richer feedback, we conduct the tests in French, as all participants have at least a B2 level in this language.

5 Results and discussion

We use the verbal data collected through think-aloud protocols and interviews to explain participants' performance and their questionnaire scores. The cross-analysis of the questionnaire scores and the interviews provide a comprehensive overview of the participant experience and highlight key patterns in participant perception. By integrating quantitative and qualitative findings, we can gain deeper insights into the strengths and weaknesses of the system and better understand participant expectations.

5.1 Performances

To evaluate the complexity and efficiency of different scenarios, we analyzed three performance metrics: completion time, completion rate, and help requests (Table 2). These metrics provide insights into participant performance, difficulty levels, and usability issues encountered during scenario execution. Completion time indicates the average duration required to complete a scenario (mean) and the variability among participants (standard deviation). A higher mean suggests a more time-consuming scenario, while a higher standard deviation reflects inconsistent performance across participants. Success rate represents the proportion of successful attempts per scenario, with 1.0 indicating a 100% success rate. Help requests (N) metric quantifies the number of instances where participants required assistance to complete the scenario, highlighting usability challenges.

Table 2. Participants' performance metrics per scenario: average completion time ($Time_{avg}$) and time variability ($Time_{SD}$) expressed in seconds, success rate, and number of help requests.

Scenario	$Time_{avg}$ (s)	$Time_{SD}$ (s)	Success rate	Help requests
S1	27.0	22.13	1.0	0
S2	73.5	47.86	1.0	2
S3	58.3	40.26	1.0	0
S4	103.4	69.79	1.0	2
S5	75.5	74.57	1.0	0
S6	125.7	69.03	0.9	6
S7	33.0	15.24	1.0	0

Among all scenarios, S6 emerges as the most difficult, exhibiting the longest completion time (125.7s), the highest variability (69.03s), and the lowest success rate (0.9). Furthermore, it records the highest number of help requests (6), indicating that participants frequently struggled. Interview data suggests that participants find this scenario cognitively demanding due to unclear guidance and cognitively challenging tasks, such as information overload or poorly adapted color schemes, which often lead them to seek external assistance.

Similarly, S4 and S2 also present usability issues, with S4 requiring considerable time (103.4s) and two help requests, suggesting that participants face obstacles but were ultimately able to complete the scenario. S2, although moderately time-consuming (73.5s), exhibits similar challenges, likely due to confusion in interface design rather than the inherent complexity of the task. Comments on S4 emphasize similar issues related to navigation and visual clarity.

5.2 UEQ+

The analysis and interpretation of UEQ+ results follow the guidelines established in the UEQ+ Handbook, ensuring scientifically robust and contextually relevant findings. Based on the collected data, Figure 4 presents mean scores and perceived importance of each scale.

Benchmark. Based on the 2023 UEQ reference dataset (452 product evaluations, 20,190 participants), each score can be benchmarked against the reference values established for the original UEQ scales [39]. As can be seen from Table 3, the solution is rated Excellent in Perspicuity, Above Average in Stimulation, and Good in Efficiency and Dependability. However, Attractiveness remains Average, and Novelty is Below Average. The Trustworthiness of Content and Quality of Content scales are not included here, as no benchmark is provided for them.

Table 3. Benchmark of our solution based on the 2023 UEQ+ reference dataset. Scores are compared against the established benchmarks for the original UEQ scales [39].

Dimension	Mean score	Benchmark
Attractiveness	1.83	Average
Perspicuity	2.68	Excellent
Efficiency	2.40	Good
Dependability	2.25	Good
Stimulation	2.25	Above Average
Novelty	0.68	Below Average

Mean scores. The Scale Means graph (Figure 4) displays the mean scores for each UX scale assessed in the UEQ+ questionnaire, providing a structured evaluation of participant perception (3). The scale ranges from -3 (extremely negative perception) to +3 (extremely positive perception), with error bars representing 95% confidence intervals, indicating the reliability and variability of responses.

Pragmatic scales show high scores. Perspicuity receives the highest rating (2.68), indicating that participants find the system particularly clear and easy to learn. Its low variance (0.28) demonstrates a strong consensus on this ease of use. Efficiency (2.40) is also well-rated, reflecting smooth task execution, although its higher variance (0.76) suggests varying experiences, possibly due to differences participant expertise. Dependability (2.25) reflects a positive perception of the system's reliability and control over interaction, though its variance (0.81) reveals that some participants encountered inconsistencies.

Hedonic scales yield more nuanced results. Stimulation (2.25) measures whether the system is exciting and motivating, with a high variance (1.01)

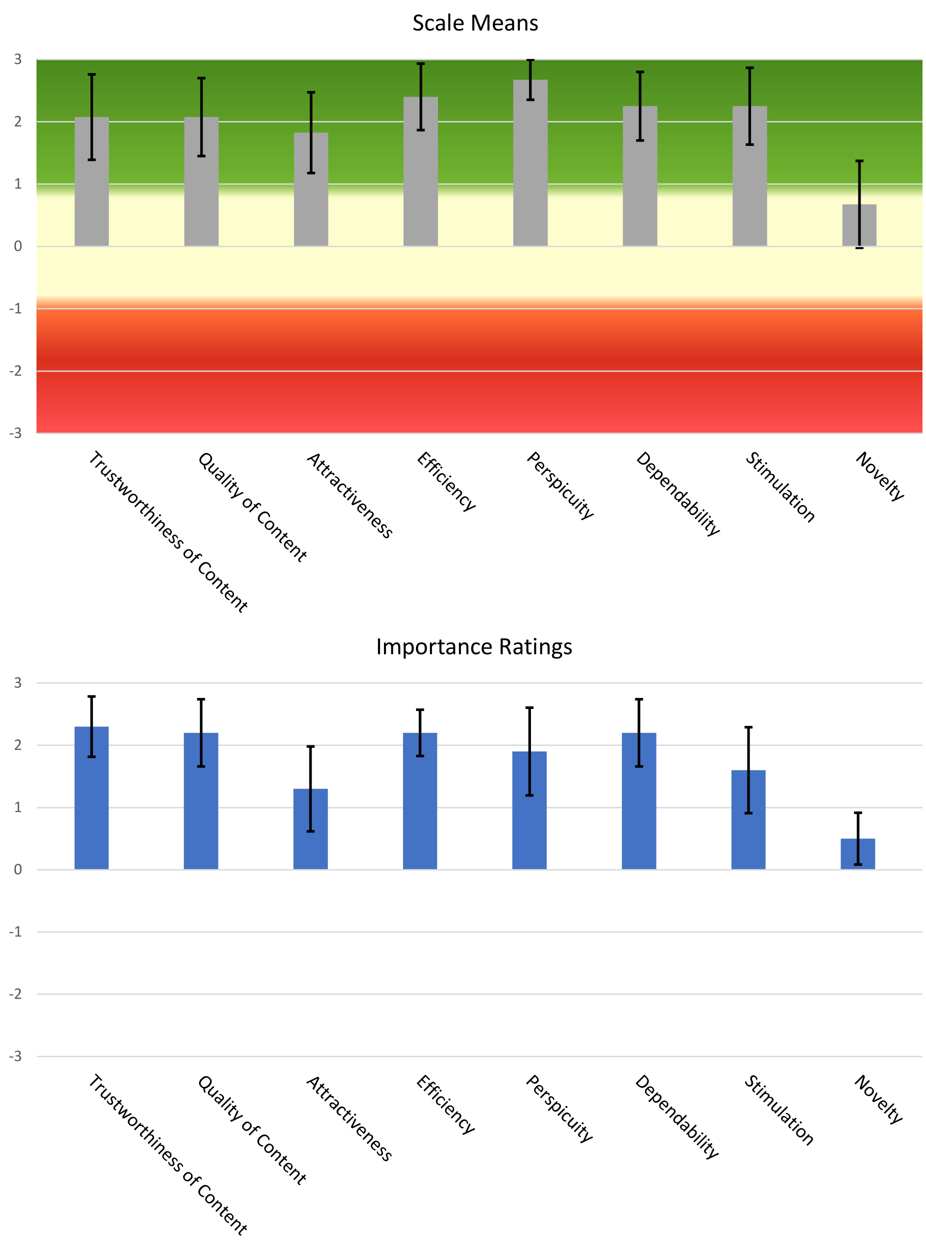

Fig. 4. UEQ+ Scale Means (top) and UEQ+ Importance Ratings (bottom)

indicating significant divergence in participant perceptions: 40% find it engaging and valuable, while others remain indifferent. Novelty (0.68)receives the lowest score, suggesting that while the system is functional, it is not widely perceived as innovative. Its high variance (1.30) and confidence interval [-0.02, 1.37] reflect strong differences in opinion: 70% recognize inventive elements, whereas others consider the system conventional. This trend may be influenced by participants with IT backgrounds, who often prioritize functionality and reliability over aesthetic appeal or innovation. Although hedonic attributes contribute to the overall experience, they are not decisive factors in participant satisfaction.

Content-related scales show a generally positive evaluation. Trustworthiness of Content and Quality of Content receive similar scores (2.08), suggesting that participants generally trust the information and find it relevant. However, Trustworthiness of Content has a higher variance (1.25) compared to Quality of Content (1.05), indicating greater divergence in opinions regarding the reliability of the information. These results suggest that, while the content is generally seen as plausible and interesting, some participants remain skeptical about its accuracy. This result can be explained by the medium fidelity of the data model, whereas the prototype is high fidelity in visual refinement, breadth of functionality, depth of functionality, and interactivity. Fidelity is a multi-dimensional concept, and mixed-fidelity prototypes can lead to varied participant perceptions [27]. The lower Trustworthiness of Content score likely stems from the data model's limited realism compared to the other high-fidelity aspects.

Relative importance. The Importance Ratings graph (Figure 4) shows that not all UX dimensions are perceived as equally important. Content-related and pragmatic scales emerge as the most critical, while hedonic aspects play a secondary role. Participants primarily prioritize content trustworthiness and usability, whereas aesthetics and novelty are less significant.

Among the highest-rated dimensions, Trustworthiness of Content (2.30) and Quality of Content (2.20) stand out, emphasizing the importance of information reliability and accuracy. This is consistent with participants' satisfaction with content-related aspects, and aligns with high mean scores in the Scale Means graph and with previous work, which underscores the role of information reliability and quality in effective online learning [45]. Efficiency (2.20), Perspicuity (1.90), and Dependability (2.20) also receive strong ratings, highlighting the importance of ease of use, task execution, and system reliability.

In contrast, hedonic scales are perceived as less critical. Stimulation (1.60) and Attractiveness (1.30) hold some relevance but are not top priorities. However, interviews provide nuanced insights that are not fully captured in the questionnaire. While Attractiveness scored moderately, 30% indicate that a more polished design could enhance their engagement. This suggests that although Attractiveness is not a primary concern, it still plays a role in influencing long-term participant adoption. Novelty (0.50) ranks as the least important UX dimension.

5.3 Ad-hoc questionnaire

We use divergent stacked bar charts to visualize the distribution of scores from the ad-hoc questionnaire, ensuring that scores of 4 and 5 indicate positive responses. Positive answers appear on the right in blue, negative ones on the left in red, and neutral responses are evenly split around the centerline in grey. Figure 5 presents the scores in descending order for gamification, recommender system, chatbot, and their integration.

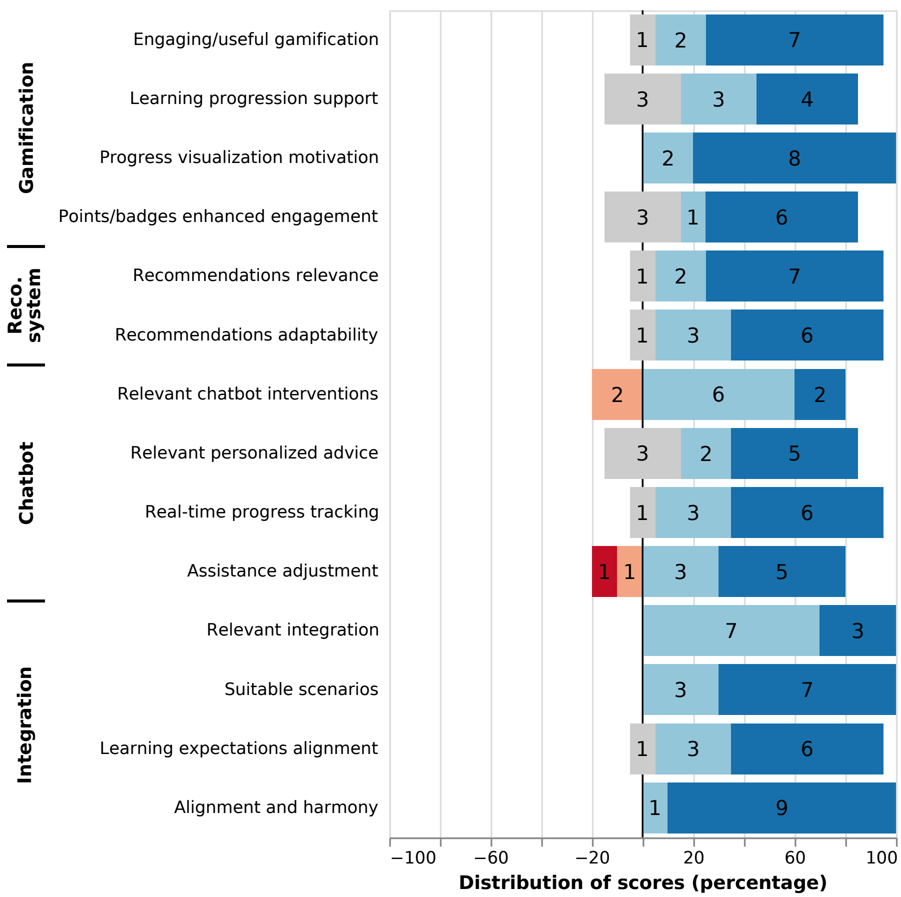

Fig. 5. Divergent stacked bar chart displaying the distribution of scores from the ad-hoc questionnaire, ordered in descending fashion for gamification, the recommender system (Reco. system), the chatbot, and their integration.

Gamification. Participants are generally satisfied with the gamification component, with scores ranging from 3 to 5. All find progress visualization motivating, 90% consider gamification engaging and useful, 70% believe these elements support a progressive learning path, and 70% say points and badges enhance engagement. Verbatim feedback confirms these results, with no negative comments received.

Recommender. The recommender system was also well-received, with scores from 3 to 5. 90% of participants find recommendations relevant and well-adapted to their progress and difficulties, a result confirmed by verbatim feedback.

Chatbot. The chatbot is the only component to receive the full range of scores, from 1 to 5, indicating mixed opinions. While 90% of participants value its real-time progress tracking and 80% find its interventions relevant and well-adapted to their performance, opinions on its personalized advice are more divided. Specifically, 70% consider the advice relevant, while others remain neutral or dissatisfied. Participants attributed their skepticism to previous experiences with generic chatbots. However, after explaining during the interview that our chatbot provides personalized assistance, all participants indicated that they would be more inclined to use it.

Integration. The integration of the three components received highly positive feedback. All participants find it relevant and find the scenarios well suited to their needs and objectives based on their role in the participant test. Additionally, 90% believe the system meets their expectations in terms of learning and objectives. Finally, no participant reported any disconnect or lack of harmony between the components.

5.4 User Goals

The results of user testing emphasize clear priorities for improving the platform. Participants primarily seek a more personalized and adaptive experience. Many express the need for scenario recommendations that align with their skill levels and progress (R3), emphasizing the importance of a recommendation system that evolves based on their interactions and performance. Additionally, participants value timely and contextual assistance from the chatbot (R8), desiring hints that support their learning process without disrupting their autonomy. Some participants also highlight the importance of a chatbot that offers guidance across multiple sessions, adapting its assistance to their evolving goals and progress (R9).

6 Conclusion and future work

This paper demonstrates the potential of a system that integrates gamification, a recommendation system, and a chatbot to improve cybersecurity training. User testing confirms that this integration effectively fosters engagement and motivation. However, feedback reveals several areas that require refinement. Participants report difficulties navigating the platform, particularly when managing preferences and goals. Additionally, the readability of information needs improvement to ensure clearer and more accessible content. Lastly, the chatbot's integration should be optimized to provide meaningful support without becoming intrusive or disruptive during the learning process.

Furthermore, it is important to test the system with a broader audience, including non-IT professionals, to evaluate its accessibility and effectiveness, as cyberattacks can affect anyone regardless of their technical background. For example, scenarios focusing on phishing awareness or secure password management are essential for making cybersecurity awareness and training more inclusive and applicable to real-world needs.

In future iterations, we aim to test the remaining requirements (R3, R8, and R9) to improve the system's overall effectiveness. We will first assess the adaptability of the recommendation system (R3) to determine how well it personalizes user experience by dynamically adjusting scenarios based on real-time interactions and progress. Next, we will evaluate the impact of contextual chatbot assistance (R8) to ensure that training hints strike the right balance between guidance and problem-solving, supporting users without reducing their learning autonomy. Finally, we will validate the chatbot's ability to adjust the narrative structure (R9) through multi-step scenario testing, assessing its effectiveness in guiding users across sessions and adapting to evolving objectives.

Acknowledgments. We would like to thank the ten participants taking part in our user test. This work was supported by the CyberExcellence project (grant number 2110186) and the CRS2 project (grant number 2110064).

References

1. Alias, S., Sainin, M.S., Fun, T.S., Daut, N.: Identification of Conversational Intent Pattern Using Pattern-Growth Technique for Academic Chatbot (1 2019)
2. Aster, A., Laupichler, M.C., Zimmer, S., Raupach, T.: Game design elements of serious games in the education of medical and healthcare professions: a mixed-methods systematic review of underlying theories and teaching effectiveness. Advances in Health Sciences Education (4 2024)
3. Barata, G., Gama, S., Jorge, J., Goncalves, D.: Engaging engineering students with gamification. In: 2013 5th International Conference on Games and Virtual Worlds for Serious Applications. pp. 1–8 (2013)
4. Barata, G., Gama, S., Jorge, J., Gonçalves, D.: Gamification for smarter learning: tales from the trenches. Smart Learning Environments **2**(1) (5 2015)
5. Chun Ho, 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 (2018)
6. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: Defining gamification. Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments pp. 9–15 (2011)

7. Durall, E., Kapros, E.: Co-design for a Competency Self-assessment Chatbot and Survey in Science Education (1 2020)
8. Ecso, A.: Understanding cyber ranges: From hype to reality (2020), https://api.semanticscholar.org/CorpusID:220626114
9. Fariani, R.I., Junus, K., Santoso, H.B.: A systematic literature review on personalised learning in the higher education context. Technology, Knowledge and Learning **28**(2), 449–476 (2023)
10. Graesser, A.C.: Conversations with AutoTutor Help Students Learn. International Journal of Artificial Intelligence in Education **26**(1), 124–132 (1 2016)
11. Haddioui, I.E., Khaldi, M.: Learner Behavior Analysis on an Online Learning Platform. International Journal of Emerging Technologies in Learning (iJET) **7**(2), 22 (6 2012)
12. 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)
13. Huang, W.Y., Soman, Y.L., Ho, S.C.: A study of gamification in e-learning design: Using a virtual reality environment. Smart Learning Environments **6**(1), 1–14 (2019)
14. Jain, S.: 160 Cybersecurity Statistics: Updated Report 2024 (11 2024), https://www.getastra.com/blog/security-audit/cyber-security-statistics/
15. Kapp, K.M.: The gamification of learning and instruction: Game-based methods and strategies for training and education. Pfeiffer (2012)
16. Kapp, K.: The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education. Business book summary, Wiley (2012)
17. Kenneally, E., Dittrich, D.: The menlo report: Ethical principles guiding information and communication technology research (2012)
18. Kieffer, S., Rukonić, L., Kervyn de Meerendré, V., Vanderdonckt, J.: A process reference model for ux. In: International Joint Conference on Computer Vision, Imaging and Computer Graphics. pp. 128–152. Springer (2019)
19. Kikalishvili, S.: Unlocking the potential of GPT-3 in education: opportunities, limitations, and recommendations for effective integration. Interactive Learning Environments **32**(9), 5587–5599 (6 2023)
20. Kuhail, M.A., Alturki, N., Alramlawi, S., Alhejori, K.: Interacting with educational chatbots: A systematic review. Education and Information Technologies **28**(1), 973–1018 (7 2022)
21. Kumar, M.N., Chandar, P.L., Prasad, A.V., Sumangali, K.: Android based educational chatbot for visually impaired people. In: 2016 IEEE international conference on computational intelligence and computing research. pp. 1–4. IEEE (2016)
22. Labadze, L., Grigolia, M., Machaidze, L.: Role of ai chatbots in education: systematic literature review. International Journal of Educational Technology in Higher Education **20** (10 2023)
23. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. vol. 5298, pp. 63–76 (11 2008)
24. Lo, C.K., Hew, K.: A comparison of flipped learning with gamification, traditional learning, and online independent study: The effects on students' mathematics achievement and cognitive engagement. Interactive Learning Environments **28**, 464–481 (05 2020)
25. Lü, L., Medo, M., Yeung, C.H., Zhang, Y.C., Zhang, Z.K., Zhou, T.: Recommender systems. Physics Reports **519**(1), 1–49 (3 2012)

26. Manzano-León, A., Camacho-Lazarraga, P., Guerrero, M.A., Guerrero-Puerta, L., Aguilar-Parra, J.M., Trigueros, R., Alias, A.: Between Level Up and Game Over: A Systematic Literature Review of Gamification in Education. Sustainability **13**(4), 2247 (2 2021)
27. McCurdy, M., Connors, C., Pyrzak, G., Kanefsky, B., Vera, A.: Breaking the fidelity barrier: an examination of our current characterization of prototypes and an example of a mixed-fidelity success. In: Proceedings of the SIGCHI conference on Human Factors in computing systems. pp. 1233–1242 (2006)
28. Mikic-Fonte, F.A., Llamas-Nistal, M., Caeiro-Rodriguez, M.: Using a Chatterbot as a FAQ Assistant in a Course about Computers Architecture. 2021 IEEE Frontiers in Education Conference pp. 1–4 (10 2018)
29. Mor, E., Santanach, F., Tesconi, S., Casado, C.: CodeLab: Designing a Conversation-Based Educational Tool for Learning to Code (1 2018)
30. Ndukwe, I.G., Daniel, B.K., Amadi, C.E.: A Machine Learning Grading System Using Chatbots (1 2019)
31. Nguyen, H.D., Pham, V.T., Tran, D.A., Le, T.T.: Intelligent tutoring chatbot for solving mathematical problems in high-school. In: 2019 11th International Conference on Knowledge and Systems Engineering. pp. 1–6 (2019)
32. Okonkwo, C.W., Ade-Ibijola, A.: Chatbots applications in education: A systematic review. Computers and Education Artificial Intelligence **2**, 100033 (1 2021)
33. Pavlov, A., Duhon, G., Dawes, J.: Examining the Impact of Task Difficulty on Student Engagement and Learning Rates. Journal of Behavioral Education **32**(3), 527–542 (11 2021)
34. Pérez, J.Q., Daradoumis, T., Puig, J.M.M.: Rediscovering the use of chatbots in education: A systematic literature review. Computer Applications in Engineering Education **28**(6), 1549–1565 (9 2020)
35. Ranoliya, B.R., Raghuwanshi, N., Singh, S.: Chatbot for university related FAQs (2017)
36. Reyssier, S., Hallifax, S., Serna, A., Marty, J.C., Simonian, S., Lavoue, E.: The Impact of Game Elements on Learner Motivation: Influence of Initial Motivation and Player Profile. IEEE Transactions on Learning Technologies **15**(1), 42–54 (2 2022)
37. Rodrigues, L., Toda, A., Palomino, P., Oliveira, W., Isotani, S.: Personalized gamification: A literature review of outcomes, experiments, and approaches (2020)
38. Schiff, D.: Out of the laboratory and into the classroom: the future of artificial intelligence in education. AI Society **36**(1), 331–348 (8 2020)
39. Schrepp, M., Hinderks, A., Thomaschewski, J.: Construction of a benchmark for the user experience questionnaire (ueq). International Journal of Interactive Multimedia and Artificial Intelligence **4**(4), 40–44 (2017)
40. Schrepp, M., Thomaschewski, J.: Design and validation of a framework for the creation of user experience questionnaires. International Journal of Interactive Multimedia and Artificial Intelligence **5**(7), 88–95 (2019)
41. Shi, L., Cristea, A.I.: Motivational Gamification Strategies Rooted in Self-Determination Theory for Social Adaptive E-Learning (1 2016)
42. Shi, L., Cristea, A.I., Hadzidedic, S., Dervishalidovic, N.: Contextual gamification of social interaction – towards increasing motivation in social e-learning. In: Advances in Web-Based Learning. pp. 116–122 (2014)
43. Sreelakshmi, A.S., Abhinaya, S.B., Nair, A., Nirmala, S.J.: A Question Answering and Quiz Generation Chatbot for Education (2019)

44. Ureta, J., Rivera, J.P.: Using chatbots to teach stem related research concepts to high school students. In: ICCE 2018—26th International Conference on Computers in Education, Work-in-Progress Poster Proceedings. pp. 338–343 (2018)
45. Vallée, A., Blacher, J., Cariou, A., Sorbets, E.: Blended learning compared to traditional learning in medical education: Systematic review and meta-analysis. Journal of Medical Internet Research **22**(8) (2020)
46. Williot, L.: Gamification and AI guidance for training in Cyber ranges. Master's thesis, Université catholique de Louvain, Belgium (2024)
47. Xiao, Y., Hew, K.F.: Personalized gamification versus one-size-fits-all gamification in fully online learning: Effects on student motivational, behavioral and cognitive outcomes. Learning and Individual Differences **113**, 102470 (2024)
48. Yuan, J., Peng, X., Liu, Y., Wang, Q.: "Don't Look at Me!": The Role of Avatars' Presentation Style and Gaze Direction in Social Chatbot Design. Computers in Human Behavior **164**, 108501 (11 2024)
49. Zhang, Q., Lu, J., Zhang, G.: Recommender systems in e-learning. Journal of Smart Environments and Green Computing **1** (04 2021)