Acta Informatica Pragensia
2026, Volume 15, Issue 3, 698–718
Special Issue: Selected Extended Papers from International Conferences CITA, ICRAMI, and ADACIS
https://doi.org/10.18267/j.aip.311

PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

Review Open Access

Gamification Meets Large Language Models: A Systematic Review of Applications and Challenges

Nurfauza Jali [1], **Lleyton Geboh Leslie** [1], **Cheah Wai Shiang** [1], **Sadok Ben Yahia** [2], **Syahrul Nizam Junaini** [1]

[1] Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia
[2] The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Denmark

Corresponding author: Nurfauza Jali (jnurfauza@unimas.my)

Editorial Record

First submission received:
October 29, 2025

Revisions received:
January 22, 2026
March 7, 2026

Accepted for publication:
March 9, 2026

Special Issue Editors:
Stephanie Chua
Universiti Malaysia Sarawak, Malaysia
Phei Chin Lim
Universiti Malaysia Sarawak, Malaysia
Siaw-Hong Liew
Universiti Malaysia Sarawak, Malaysia
Kim-Mey Chew
Universiti Malaysia Sarawak, Malaysia

Academic Editor:
Zdenek Smutny
Prague University of Economics and Business, Czech Republic

This article was accepted for publication by the Academic Editor upon evaluation of the reviewers' comments.

How to cite this article:
Jali, N., Leslie, L. G., Shiang, C. W., Yahia, B. S., & Junaini, S. N. (2026). Gamification Meets Large Language Models: A Systematic Review of Applications and Challenges. *Acta Informatica Pragensia*, 15(3), 698–718. https://doi.org/10.18267/j.aip.311

Abstract

Background: The convergence of large language models (LLMs) and gamification has emerged as a promising direction for enhancing user engagement, personalisation and interaction quality across digital systems. While prior studies have largely examined LLMs or gamification in isolation, limited work has systematically explored their combined role in redefining engagement paradigms beyond static reward-based mechanisms.

Objective: This review systematically examines how LLMs are integrated into gamified systems, identifies their application contexts and functional roles and synthesises reported benefits, challenges and methodological trends across existing empirical studies.

Methods: This systematic literature review examines peer-reviewed studies published between 2020 and 2025, following established systematic review protocols and the PRISMA 2020 framework. A total of 718 articles were retrieved from Scopus, IEEE Xplore, ScienceDirect and the ACM Digital Library. After applying predefined inclusion and exclusion criteria, 27 studies were selected for in-depth analysis. A quality assessment process was conducted to evaluate methodological rigour, consistency and reporting clarity, ensuring the reliability of the synthesis.

Results: The review identifies key application domains where LLMs are integrated into gamified systems, including intelligent non-player characters, adaptive educational feedback, persuasive gamified environments and embodied digital avatars for social and behavioural simulations. Across these contexts, LLMs were found to substantially enhance personalisation, narrative immersion and learner motivation, complementing the motivational structures of gamification. However, recurring challenges were also reported, including hallucination risks, ethical concerns in social and extended reality contexts and misalignment between LLM-driven system behaviour and user expectations.

Conclusion: This review provides a comprehensive synthesis of the opportunities, limitations and methodological trends associated with integrating LLMs into gamified environments. By consolidating evidence across diverse application domains, it offers practical and theoretical guidance for researchers, educators and system designers on responsible and effective use of LLMs to support adaptive, engaging and ethically grounded gamified systems.

Index Terms

Adaptive learning; Game design; Interactive narratives; LLM; User engagement.

1 INTRODUCTION

The integration of large language models (LLMs) into gamified systems is emerging as a transformative trend within human-computer interaction, particularly in fields such as education, simulation and digital engagement.

LLMs such as GPT-4 can generate coherent content and simulate human-like behaviour, enabling the creation of more interactive and personalised experiences. These models enhance conventional gamification, which often relies on fixed-point reward systems, by transforming them into dynamic, adaptable frameworks that can intelligently respond to user input.

Gamified environments refer to the application of game design elements in non-game contexts to enhance user engagement, motivation and learning outcomes (Deterding et al., 2011; Seaborn & Fels, 2015). Deterding et al. (2011) defined gamification as "the use of game design elements in non-game contexts", a concept now widely adopted in digital learning and HCI research. Recent advances in artificial intelligence, particularly large language models (LLMs), have created new opportunities for personalisation, dynamic narrative and adaptive feedback in these environments (Brown et al., 2020).

Building upon these developments, the authors previously presented a systematic literature review at the 14th International Conference on Information Technology in Asia (CITA 2025), which explored the integration of LLMs in gamified systems (Leslie et al., 2025). That study found that most existing research focused on educational and simulation-based applications in which LLMs acted as intelligent agents capable of adapting to learner behaviour and providing personalised feedback. It also highlighted key challenges such as content hallucination, bias and the lack of standardised evaluation metrics when applying LLMs in gamified contexts. However, as a conference-length paper, it offered only a preliminary synthesis of findings based on a limited dataset.

Recent studies have demonstrated how LLMs can serve as dynamic agents in diverse environments, including learning platforms, virtual social spaces and experience-driven simulations (Sissler, 2024; Khanshan et al., 2024). A prominent application involves enhancing non-player characters (NPCs) in game engines such as Unity. Rather than relying on pre-scripted responses, LLM-powered NPCs can engage in real-time, context-aware conversations, thereby improving immersion and responsiveness in gameplay scenarios (Sissler, 2024).

In retail, for instance, LLMs support the gamification of shopping through persuasive design strategies (Ataguba et al., 2025). In software testing and simulation, LLMs generate realistic user behaviour to validate system robustness (Khanshan et al., 2024). Meanwhile, in corporate and educational settings, LLM-driven digital avatars or "virtual twins" participate in meetings on behalf of users, incorporating gamified features such as trust and performance metrics (Leong et al., 2024).

Despite their potential, LLMs also pose challenges. Issues such as bias, hallucinations and inconsistencies can affect both user experience and ethical deployment. Therefore, integrating LLMs into gamified systems requires rigorous design consideration to ensure responsible behaviour, transparency and user trust.

This systematic analysis examines 27 selected experiments, focusing on how LLMs enhance key gamification components, including adaptive narratives, user behaviour modelling, intelligent NPCs and social avatars. The results offer significant insights into the expanding influence of LLMs in transforming gamified interactions across various areas, including education, software systems and digital communication.

2 RELATED WORKS

Interest in integrating artificial intelligence (AI) into gamified systems has grown substantially, particularly with the advancement of LLMs. Unlike earlier systems that relied on static dialogue and predetermined behaviour, LLMs enable more fluid, context-aware and human-like interactions, significantly enhancing user engagement (Sissler, 2024). These capabilities mark a shift from rigid design structures towards systems that can adapt dynamically to user input.

A key improvement brought about by LLMs is in behavioural modelling. Khanshan et al. (2024) addressed the limitations of traditional simulation approaches, which required extensive manual configuration. They introduced Prompt4Code, a framework that makes use of LLMs to simulate nuanced human behaviour within gamified environments, thereby reducing development overhead while increasing behavioural complexity and realism.

In the domain of social interaction, Leong et al. (2024) proposed Dittos, intelligent digital avatars designed to emulate human-like behaviour in collaborative scenarios. These LLM-powered agents can engage in natural dialogue, foster trust and participate in activities such as virtual meetings, demonstrating the applicability of LLMs across both social and gamified systems.

Furthermore, the role of LLMs is evolving from being mere supportive tools to becoming adaptive agents that enhance personalisation, interactivity and user immersion. A recent survey by J. Lin et al. (2025) reinforces this shift, highlighting the growing potential of LLMs to enhance recommender systems, with applications that closely align with gamified recommendation strategies. Table 1 summarises related work that underpins the findings and insights presented in this review.

Table 1. *Summary of related work aligned with research goals.*

Reference	Goals based on research questions
Sissler (2024)	• Investigate the use of GPT-3.5 in enhancing non-player character (NPC) interactions in Unity. • Explore how LLMs improve immersion through natural, unscripted dialogue.
Khanshan et al. (2024)	• Introduce Prompt4Code to simulate participant behaviour using LLMs. • Demonstrate the adaptability of LLMs in experience sampling and gamified simulation contexts.
Leong et al. (2024)	• Present "Dittos", embodied agents for user representation in collaborative settings. • Explore gamified dynamics, such as trust and social intelligence, through LLM agents.
J. Lin et al. (2025)	• Examine how LLMs can enhance recommender systems.

3 REVIEW METHOD

This section presents the structured methodology used to conduct a systematic literature review (SLR) on the application of LLMs in gamified systems. The review process was designed to be transparent, replicable and comprehensive. It encompasses review planning, literature search and selection, followed by data extraction and analysis.

3.1 Review planning

The primary objective of this review is to explore how LLMs are utilised within gamified contexts across various domains, including education, simulation and human-computer interaction. To achieve this, a structured approach was adopted involving the formulation of precise research questions and objectives to guide the study.

3.1.1 Research questions

The following research questions were developed to frame the review and ensure alignment with the overall objectives:

- RQ1: How are LLMs used to support or enhance gamified systems?
- RQ2: What gamification mechanics or applications are commonly integrated with LLMs?
- RQ3: What benefits and challenges are reported in the literature regarding LLM integration into gamified environments?

These questions aim to assess the current research landscape, highlight practical implementations and identify existing knowledge gaps at the intersection of LLMs and gamification.

3.1.2 Search strategy

A well-defined search strategy was employed to identify and systematically capture relevant, high-quality studies. Following the PRISMA 2020 guidelines for transparent reporting of systematic reviews (Page et al., 2021), the process consisted of four stages: identification, screening, eligibility and inclusion.

3.1.3 Source selection

To ensure comprehensive and credible coverage of the literature, four leading academic databases were selected: Scopus, ScienceDirect, IEEE Xplore and the ACM Digital Library. These platforms were chosen for their extensive coverage of peer-reviewed research in artificial intelligence, gamification, educational technology and human-computer interaction.

3.1.4 Keyword/search term selection

A consistent Boolean query was employed across all the selected databases:

```
("Large Language Models" OR "LLM") AND "Gamification"
```

Additional terms such as "game-based learning", "simulation", "NPCs" and "adaptive tutoring" were used iteratively to refine results. The search period covered January 2020 – April 2025, reflecting the rapid adoption of LLMs during this time.

3.1.5 Inclusion and exclusion criteria

As the literature is collected, it needs to be assessed to ensure its usefulness for the review. There are two types of criteria used to select relevant literature. The criteria are grouped into inclusion and exclusion criteria. These criteria are used to filter the literature found, ensuring that it can be used as primary studies in this review. For inclusion criteria, a set of conditions has been identified that will be included in the review, while exclusion criteria are a set of conditions that will exclude literature from the review. The inclusion criteria required that studies (i) involve the use of LLMs in gamified or game-inspired environments, (ii) present empirical or design-focused contributions and (iii) be published in peer-reviewed journals. Exclusion criteria removed grey literature, preprints, position papers and studies without a clear gamification context. A summary is shown in Table 2.

Table 2. *Inclusion and exclusion criteria.*

Inclusion criteria	Exclusion criteria
The publication date is between January 2020 and May 2025.	The publication date is outside the range 2020–2025.
The full-text of the article is available.	The full text is inaccessible or behind a paywall.
The article addresses the integration of LLMs with gamification elements – RQ1, RQ2, RQ3.	The article does not address LLMs or gamification.
The article is published in a peer-reviewed journal.	The article is a conference paper, preprint, book chapter or any non-journal publication.
The article discusses the application of LLMs in education, simulation, storytelling and interactive environments.	The article focuses solely on the technical model performance, without providing context for user or system integration.

3.2 Conducting the review

The review process began with a comprehensive literature search across four primary academic databases: Scopus, IEEE Xplore, ScienceDirect and the ACM Digital Library. These sources were selected due to their strong coverage of research into artificial intelligence, gamification and human–computer interaction, as well as their robust filtering and indexing capabilities. Searches were conducted using combinations of keywords applied to article titles, abstracts, metadata and keywords.

Although additional databases, such as Web of Science (WoS), Taylor & Francis and SpringerLink, were initially considered, they were excluded from the final analysis due to the limited number of unique results or significant overlap with the primary databases.

3.2.1 Search and study selection

The search was conducted using a predefined query presented in Section 3.1.4. This query was consistently applied across all four selected databases to ensure a precise and replicable retrieval process. The search yielded a total of 718 records, distributed as follows:

- ACM Digital Library: 605
- ScienceDirect: 106
- Scopus: 7
- IEEE Xplore: 0

After removing duplicates and an initial screening of titles and abstracts, 48 studies were shortlisted for full-text evaluation. Each article was then assessed against the inclusion and exclusion criteria established during the review planning phase.

Of these, 21 articles were excluded for various reasons, including a lack of direct relevance to gamification, the absence of applied LLM integration or a primary focus on technical model design rather than user-centred systems. As a result, 27 studies were selected for final inclusion and detailed analysis in this review.

3.2.2 Data extraction and synthesis

For each of the 27 selected studies, relevant data were systematically extracted to ensure consistency and comparability. Key elements recorded included the author(s), year of publication, application domain (e.g., education, simulation, narrative), mode of LLM integration (such as adaptive feedback, personalised progression or role-based interaction) and any documented outcomes, benefits or implementation challenges. This structured approach enabled a comprehensive synthesis of how LLMs are applied within gamified environments across various contexts.

3.2.3 Reporting the review

The reporting of this review is based on the checklist provided by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 checklist (Page et al., 2021), see also Appendix A. The results of the review are reported in detail in Section 4. The review report includes a detailed explanation of the extracted data, findings related to the research questions and a discussion of the results. To ensure transparency and academic rigour, all the sources are correctly cited through in-text references and compiled in a complete reference list. The overall workflow is summarised in the PRISMA flow diagram (see Figure 1), which visually depicts the identification, screening, eligibility and inclusion stages.

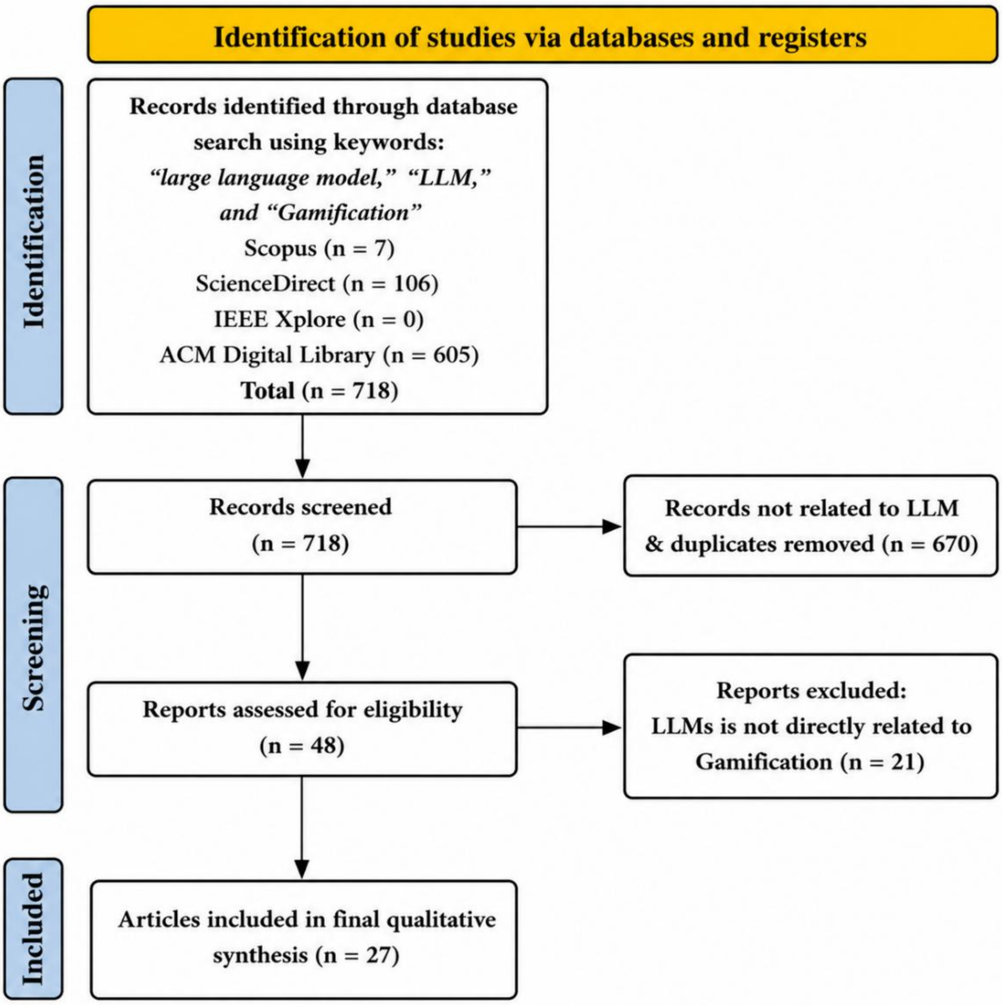

Figure 1. *PRISMA study selection process for primary studies.*

4 RESULTS

This section presents the findings of the systematic literature review in relation to the three research questions (RQ1–RQ3), based on a detailed analysis of 27 peer-reviewed articles examining the application of large language models (LLMs) in gamified systems. To support an initial overview of the conceptual landscape, a keyword co-occurrence map was generated using VOSviewer to visualise recurring terms across the selected studies.

Figure 2 presents the resulting keyword co-occurrence visualisation. Owing to the emergent and interdisciplinary nature of research at the intersection of LLMs and gamification, only a limited number of keywords consistently met the co-occurrence threshold. Consequently, the visualisation highlights several dominant conceptual anchors, including large language models, generative AI, collaboration and behaviour simulation, rather than forming dense or clearly separated thematic groupings. These recurring terms indicate sustained research interest in interactive content generation, collaborative systems and simulation-driven applications within gamified environments.

The keyword co-occurrence map is therefore intended to provide high-level thematic orientation and transparency in the review process. Detailed interpretation of research themes, applications, benefits and challenges is derived from the qualitative synthesis presented in the following subsections.

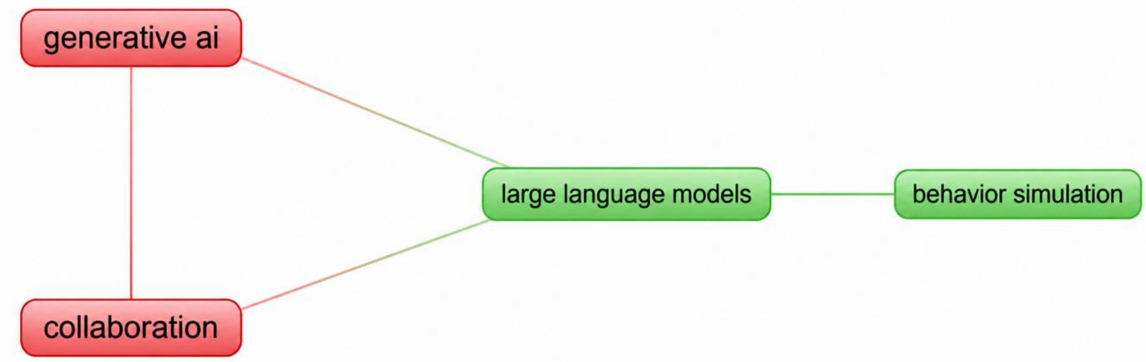

Figure 2. *Visualisation of keywords used from selected primary studies.*

4.1 How are LLMs used to support or enhance gamified systems? (RQ1)

The reviewed literature demonstrates that large language models function as core enabling components in contemporary gamified systems, supporting interaction, adaptation and behavioural realism. Rather than serving as peripheral content generators, LLMs are increasingly embedded as decision-making, conversational and simulation engines that enhance immersion, engagement and system flexibility.

One of the most common uses of LLMs is the replacement of static, rule-based dialogue systems with dynamic conversational agents. In game and simulation environments, LLM-driven non-player characters generate unscripted, context-aware responses that adapt to user actions and prior interaction history. For example, Sissler (2024) enhanced NPCs in Unity 3D using GPT-3.5, enabling continuous dialogue without predefined trees and allowing narrative flow to emerge through interaction. Extending this approach, Li et al. (2025) demonstrated LLM-based agents in a job fair simulation capable of autonomous decision-making, role negotiation and team formation, illustrating how LLMs support complex social behaviour in gamified 3D environments.

In educational gamification, LLMs are commonly deployed as adaptive tutors, co-learners or feedback agents. Rather than providing static hints or scores, these systems analyse learner input and generate personalised explanations, prompts and reflective feedback. Bany Abdelnabi et al. (2025) showed how LLMs deliver tailored instructional feedback during clerkship training, while Kumar et al. (2024) reported positive effects on student engagement and performance in supported learning environments. In these contexts, LLMs support formative assessment by transforming feedback into an interactive and motivating process.

Behaviour simulation represents another prominent application area. LLMs are used to model human-like participants, enabling the testing and evaluation of gamified systems without relying on large numbers of real users. Khanshan et al. (2024) introduced Prompt4Code, which employs in-context learning to simulate participant behaviour in experimental settings. Similarly, Wang et al. (2025) utilised LLMs to generate realistic user behaviour simulations, supporting large-scale testing and iterative design of gamified systems. These approaches highlight how LLMs contribute to scalability and experimental control.

In socially oriented gamified environments, LLMs are integrated as embodied or proxy agents that mediate interaction between users. Leong et al. (2024) proposed Dittos, LLM-driven avatars that represent users in virtual meetings and incorporate social gamification constructs such as trust, delegation and agency. In a related domain, Spitale et al. (2025) introduced VITA, a robotic coaching system that uses LLMs to deliver adaptive well-being support over time, illustrating how conversational gamification can extend beyond purely digital interfaces.

Narrative-centred applications further demonstrate the versatility of LLMs in gamification. Yan et al. (2025) presented a framework for dynamically generating branching storylines in social life simulations, where narratives evolve in response to user behaviour and dialogue. This enables personalised storytelling experiences that move beyond linear narratives and reinforce player agency.

Overall, the literature indicates that LLMs are most commonly used to support gamified systems as conversational engines, adaptive feedback providers, behaviour simulators, social agents and narrative generators. These functional roles collectively enhance interaction quality, realism and adaptability. Table 3 summarises the primary ways in which LLMs are employed across the reviewed studies in relation to RQ1.

Table 3. *Applications of LLMs in gamified systems (RQ1).*

No.	Application context	LLM role/function	Primary study
1	NPC interaction (games)	Replacing static dialogue scripts with dynamic, context-aware natural language interaction	Sissler (2024)
2	3D simulation agents	Enabling autonomous reasoning, role negotiation and decision-making in virtual environments	Li et al. (2025)
3	Educational tutoring	Providing adaptive feedback, personalised guidance and co-learning support	Bany Abdelnabi et al. (2025), Kumar et al. (2024)
4	Behaviour simulation	Simulating human participant responses for experimental testing and system evaluation	Khanshan et al. (2024)
5	Digital avatars (collaboration)	Representing users through embodied or proxy agents supporting trust, delegation and teamwork	Leong et al. (2024)
6	Narrative storytelling	Generating dynamic, branching narratives that adapt to user actions and dialogue	Yan et al. (2025)
7	Scalable user testing	Simulating diverse user behaviour to support large-scale system testing and iteration	L. Wang et al. (2025)
8	Wellness coaching simulation	Delivering adaptive, long-term coaching and feedback in gamified mental well-being contexts	Spitale et al. (2025)

4.2 What gamification mechanics or applications are commonly integrated with LLMs? (RQ2)

The reviewed literature indicates that integrating large language models enables a shift from surface-level gamification mechanics towards interaction-centric, adaptive game designs. Rather than relying primarily on points, badges or static progression systems, LLM-driven gamified environments commonly incorporate mechanics centred on dialogue, adaptation, simulation and narrative co-creation.

One of the most prevalent mechanics is branching and emergent dialogue, particularly by using LLM-powered non-player characters and social agents. Unlike traditional dialogue trees, these systems generate responses dynamically based on conversational context, user history and inferred intent. Studies such as Li et al. (2025) and Sissler (2024) show that LLM-driven dialogue supports negotiation, persuasion and role-based interaction, enabling players or learners to influence outcomes through natural language exchanges rather than predefined choices. This mechanic is frequently applied in roleplay scenarios, serious games and social simulations.

A second recurring application is adaptive feedback and personalised learning loops in educational gamification. LLMs are commonly used to analyse user input, detect misconceptions and generate context-sensitive feedback that adapts in complexity and tone. Systems such as ALure (Neshaei et al., 2025) demonstrate how learner errors are transformed into reflective prompts, hints or follow-up challenges, effectively gamifying the feedback process itself. In these contexts, progression is tied not only to task completion but also to demonstrated understanding and iterative improvement.

Simulation-based roleplay represents another dominant application area. Several studies have employed LLMs to populate simulated environments with autonomous agents capable of social reasoning, collaboration or conflict. Lin et al. (2025) illustrated this approach in SimSpark, where LLMs enable large-scale behavioural simulations without live participants. Similar mechanics appear in training, testing and rehearsal systems, where users interact with simulated stakeholders, teammates or adversaries in gamified scenarios that approximate real-world complexity.

Social gamification mechanics are also frequently integrated when LLMs are embodied as avatars or intelligent agents. In such systems, mechanics such as trust scores, relationship states, collaboration metrics and role-based achievements are layered on top of conversational interaction. Leong et al. (2024) and Wester et al. (2024) have shown how LLM-driven agents act as social counterparts rather than passive interfaces, allowing gamified systems to model cooperation, empathy or moral judgement within controlled interaction frameworks.

Finally, dynamic narrative generation and personalised storytelling emerge as a distinctive LLM-enabled mechanic. Rather than following linear storylines, LLM-driven systems generate or adapt narratives in response to player decisions, dialogue and behavioural patterns. Yan et al. (2025) demonstrated how narrative branches evolve in real time, supporting long-term engagement through personalised story arcs and evolving character relationships. This form of narrative gamification is particularly prominent in exploratory games, life simulations and creative learning environments.

Collectively, these findings indicate that LLMs are most commonly integrated into gamified systems as enablers of adaptive interaction, social simulation and context-aware feedback, rather than as replacements for conventional reward structures. Table 4 summarises the dominant gamification mechanics and application domains identified across the reviewed studies.

Table 4. *Common gamification mechanics adopted with LLMs (RQ2).*

No.	Gamification mechanic	Description / How LLMs are used	Representative studies
1	Branching and emergent dialogue	Generating non-linear, context-aware dialogue that adapts to user input, history and inferred intent rather than predefined dialogue trees	Sissler (2024); Wu et al. (2025)
2	Adaptive feedback loops	Analysing user responses and providing personalised, error-sensitive feedback, hints or reflective prompts to support progression	Neshaei et al. (2025); Bany Abdelnabi et al. (2025)
3	Personalised learning progression	Dynamically adjusting task difficulty, content sequencing and feedback depth based on individual performance and behaviour	Bany Abdelnabi et al. (2025); Wang et al. (2025)
4	Simulation-based roleplay	Populating simulated environments with autonomous agents capable of social reasoning, collaboration or conflict	Lin et al. (2025); Crandall and Crandall (2024)
5	Social gamification mechanics	Modelling trust, collaboration, agency and role-based interaction through LLM-driven avatars or agents	Leong et al. (2024); Wester et al. (2024)
6	Narrative co-creation	Dynamically generating or adapting storylines and character arcs in response to user actions and dialogue	Yan et al., (2025); Meyerson et al. (2024)
7	Behaviour-driven progression	Linking progression, rewards or outcomes to demonstrated behaviour, reasoning or interaction quality rather than simple completion	Li et al. (2025); Sissler (2024)
8	Reflective error-based mechanics	Transforming user errors or suboptimal decisions into gamified learning opportunities through guided reflection and revision	Neshaei et al. (2025); Wang et al. (2025)

Figure 3 illustrates the cycle of gamification mechanics enabled by LLMs, highlighting their roles in creating adaptive, dynamic and immersive user experiences.

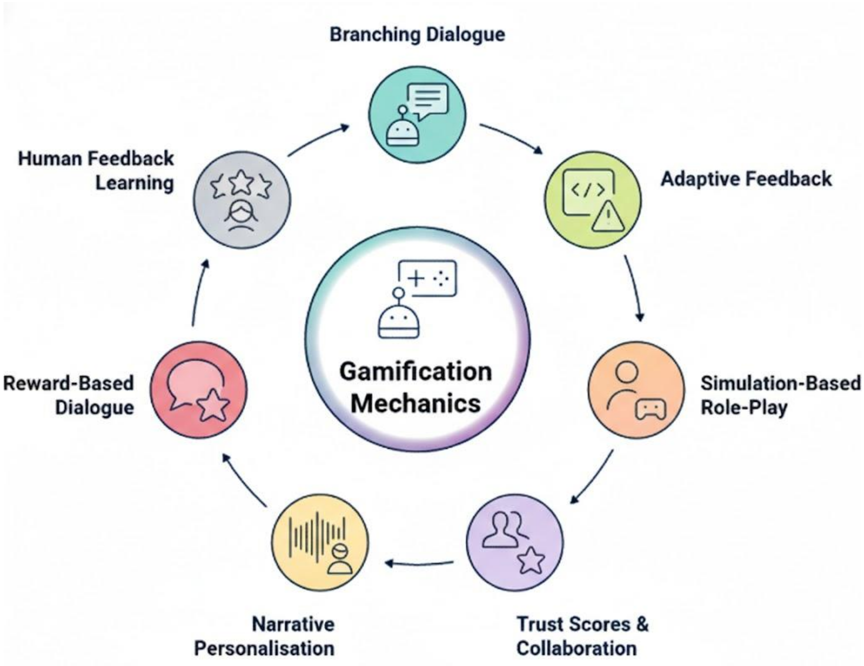

Figure 3. *LLM-enabled gamification mechanics enhancing adaptive and interactive experiences*.

4.3 What benefits and challenges are reported in the literature regarding LLM integration into gamified environments? (RQ3)

The literature reveals that integrating large language models into gamified environments introduces a combination of transformative benefits and persistent technical and ethical challenges. Across the reviewed studies, LLMs are consistently shown to enhance realism, scalability, personalisation and interaction quality. At the same time, concerns related to output reliability, ethical responsibility and alignment with system goals remain prominent.

One of the most frequently reported benefits is enhanced immersion through naturalistic and context-aware interaction. Li et al. (2025) and Ronagh Nikghalb and Cheng (2025) have demonstrated that LLM-driven dialogue enables fluid, human-like communication that sustains user engagement across both entertainment-focused and instructional gamified systems. By reducing reliance on scripted interactions, LLMs allow users to participate in open-ended exchanges that better resemble real-world social interaction, thereby strengthening perceived realism and motivation.

Scalability represents another significant advantage. Unlike traditional gamified systems that depend on predefined rules or live human facilitation, LLMs can simulate diverse behaviour at scale. Crandall and Crandall (2024) illustrated this capacity through large-scale automated software testing scenarios, where LLMs enabled repeated, varied interaction without additional human resources. Such scalability is particularly valuable in educational simulations, training environments and large multiplayer contexts.

Personalisation also emerges as a core strength of LLM integration. Bany Abdelnabi et al. (2025) reported that LLMs can tailor feedback, explanations and task complexity to individual learner profiles, thereby improving instructional relevance and engagement. Rather than offering uniform progression paths, LLM-driven systems adapt dynamically to user performance and behaviour, supporting more inclusive and responsive gamified experiences.

In collaborative and socially oriented simulations, LLMs further contribute by enabling realistic agent behaviour and decision-making. Leong et al. (2024) and Wester et al. (2024) have demonstrated how LLM-driven avatars and agents can simulate trust, cooperation and social agency, enriching roleplaying scenarios and multi-user interactions. These capabilities allow gamified systems to model complex social dynamics that are difficult to achieve with rule-based agents alone.

Despite these advantages, several recurring challenges are consistently reported. A major technical concern is content hallucination, where LLMs generate responses that appear plausible but are factually incorrect or misleading. Song et al. (2025) highlighted this issue as particularly problematic in knowledge-intensive and evaluative contexts, where accuracy is critical to learning outcomes and user trust.

Ethical risks also feature prominently in the literature. Baldry et al. (2024) discussed issues related to bias, misinformation and privacy, especially in mixed-reality and health-related gamified applications. These risks underscore the need for transparency, monitoring and ethical safeguards when deploying LLM-driven systems in sensitive domains.

From a design standpoint, alignment and control remain challenging. The generative and probabilistic nature of LLMs can produce outputs that diverge from predefined learning objectives or narrative structures. Sissler (2024) observed that such unpredictability may lead to narrative drift or inconsistency, potentially weakening instructional coherence or player experience if not carefully constrained.

Finally, trust and anthropomorphism introduce important socio-technical concerns. Users may attribute emotional intelligence, moral reasoning or authority to LLM-driven agents beyond their actual capabilities. Wester et al. (2024) cautioned that this misperception can result in misplaced reliance or emotional attachment, raising questions about responsibility, accountability and user well-being.

In summary, the literature presents a balanced view of LLM integration in gamified systems. While LLMs substantially enhance immersion, adaptability and scalability, their effective deployment requires deliberate design choices, ethical oversight and continuous evaluation to mitigate risks and preserve user trust. Table 5 summarises the key benefits and challenges identified across the reviewed studies for RQ3.

Table 5. *Reported benefits and challenges of LLM integration in gamified systems (RQ3).*

No.	Category	Benefit / challenge	Description	Representative studies
1	Benefit	Enhanced immersion	Natural, context-aware dialogue enables human-like interaction, increasing realism and sustained engagement	Li et al. (2025); Ronagh Nikghalb and Cheng (2025)
2	Benefit	Scalability	Ability to simulate diverse behaviour and interactions at scale without live human facilitation	Crandall & Crandall (2024); Wang et al. (2025)
3	Benefit	Personalisation	Adaptive feedback, content sequencing and difficulty adjustment based on individual user profiles and performance	Bany Abdelnabi et al. (2025)
4	Benefit	Social realism	Simulation of trust, collaboration and social agency through LLM-driven avatars and agents	Leong et al. (2024); Wester et al. (2024)
5	Benefit	Narrative flexibility	Dynamic and branching storytelling that evolves in response to user behaviour and dialogue	Yan et al. (2025)
6	Challenge	Content hallucination	Generation of plausible but inaccurate or misleading outputs, particularly problematic in knowledge-sensitive contexts	Song et al. (2025); Sissler (2024)
7	Challenge	Alignment and control	Difficulty ensuring that LLM outputs remain consistent with predefined learning objectives or narrative structures	Sissler (2024)
8	Challenge	Ethical risks	Risks related to bias, misinformation, privacy and inappropriate content in sensitive domains	Baldry et al. (2024)
9	Challenge	Overtrust and anthropomorphism	Users attributing emotional intelligence, moral agency or authority to LLM-driven agents, leading to misplaced reliance	Wester et al. (2024)
10	Challenge	Evaluation limitations	Limited longitudinal metrics for assessing learning outcomes, behavioural change or sustained engagement	Lin et al. (2025); Song et al. (2025)

5 DISCUSSION

This section synthesises the key findings, opportunities and limitations derived from the 27 studies reviewed on the application of large language models (LLMs) in gamified systems. The integration of LLMs is gaining momentum, particularly in domains such as storytelling, education and simulation, where adaptability and interaction are crucial. Although only a limited number of studies focus exclusively on gamification, many provide important insights into how LLMs can enhance personalisation, flexibility and user engagement.

Table 6 summarises the main benefits, challenges, gamification mechanics and roles/functions of LLMs across the reviewed studies. To improve interpretability and provide a higher-level synthesis of these findings, Figure 4 presents a thematic visual summary of the five dominant clusters identified in the literature: (i) engagement and narrative immersion, (ii) personalisation and adaptive support, (iii) collaboration and social simulation, (iv) scalability and behaviour simulation, and (v) ethics, reliability and systemic risks.

Figure 4. *Thematic clustering of benefits, challenges, gamification mechanics and LLM-driven roles identified across reviewed studies on gamified learning systems.*

5.1 Engagement and narrative immersion

As illustrated in Figure 4, engagement-driven applications constitute the most prominent cluster in the reviewed literature. Across games, simulations and educational platforms, LLMs are frequently employed as dynamic dialogue partners and narrative engines that support emergent storytelling and adaptive non-player character (NPC) interactions. Compared with traditional scripted dialogue systems, LLM-driven approaches consistently demonstrate improved immersion, perceived realism and user motivation.

However, cross-study synthesis reveals recurring challenges that temper these benefits. Several studies have reported issues related to narrative drift, where generated content deviates from intended learning objectives or story coherence. Others highlight superficial motivational effects when narrative generation is not meaningfully integrated with underlying gamification mechanics. These findings suggest that while LLMs are effective in enhancing engagement at the interaction level, their impact depends strongly on careful narrative control and alignment with instructional or gameplay goals.

5.2 Personalisation and adaptive support

Personalisation and adaptive support emerge as a second dominant theme, particularly within educational gamified systems. In this cluster, LLMs are commonly positioned as adaptive tutors capable of delivering personalised feedback, interpreting learner errors and tailoring responses to individual progress and context. Compared to rule-based adaptive systems, LLM-driven approaches offer greater flexibility and responsiveness, enabling more nuanced learner modelling.

Despite these advantages, the synthesis highlights a consistent set of risks associated with over-adaptation. Several studies have cautioned that excessive personalisation may reduce productive struggle, foster learner dependency or obscure assessment transparency. These concerns point to a critical design tension between adaptiveness and pedagogical control. Future systems must therefore balance responsiveness with structured scaffolding to ensure that personalisation supports, rather than undermines, long-term learning outcomes.

5.3 Collaboration and social simulation

The collaboration and social simulation cluster reflects growing interest in using LLMs to model peer interaction, teamwork and trust dynamics in gamified environments. Studies in this area demonstrate how LLM-driven agents can simulate collaborative partners, negotiate roles and participate in group decision-making processes, thereby enhancing social presence and realism.

At the same time, cross-study comparison reveals unresolved challenges related to anthropomorphism and overtrust. Users may attribute excessive competence, authority or intent to LLM-based agents, potentially distorting collaborative dynamics or decision-making. These findings highlight the need for transparency mechanisms that clearly communicate agent capabilities and limitations, particularly in educational and policy-oriented gamified systems.

5.4 Scalability and behaviour simulation

Scalability and behaviour simulation form a fourth cluster that emphasises the use of LLMs for large-scale user modelling, persona generation and simulation-based evaluation. These approaches are especially valuable in contexts where recruiting human participants is impractical or costly. The reviewed studies demonstrate that LLMs can efficiently generate diverse behavioural profiles and simulate interaction patterns at scale.

However, this cluster also exposes one of the most significant research gaps identified in the review. Many studies have relied on short-term, proxy or system-centric evaluation metrics that provide limited insight into behavioural validity or long-term impact. The tendency to oversimplify human behaviour raises concerns about the ecological validity of simulation results. Addressing this gap will require more rigorous evaluation designs, including longitudinal studies and mixed-method approaches that combine quantitative metrics with qualitative validation.

5.5 Ethics and reliability as cross-cutting concerns

Ethics and reliability risks emerge as a cross-cutting concern spanning all the thematic clusters. Persistent issues related to hallucination, bias, misinformation, privacy and governance constrain the feasibility of fully autonomous LLM-driven gamified systems. While many studies have acknowledged these risks, comparatively few have provided concrete mitigation strategies or formal evaluation frameworks.

The synthesis indicates broad consensus on the necessity of human-in-the-loop safeguards, transparent model behaviour and ethical governance mechanisms. These requirements are particularly critical in educational and social simulation contexts, where unintended outputs may influence learner behaviour or decision-making. Strengthening ethical alignment and reliability assessment remains a central challenge for future research.

5.6 Summary and implications for research

Overall, the combined use of Table 6 and Figure 4 enables a structured cross-study synthesis that clarifies how LLMs are currently operationalised within gamified systems. While the literature demonstrates clear benefits in engagement, personalisation, collaboration and scalability, it also reveals consistent limitations related to evaluation rigour, ethical governance and long-term impact assessment.

These findings highlight several priority directions for future work. Firstly, there is a need for longitudinal and theory-driven evaluation frameworks that move beyond short-term engagement metrics. Secondly, stronger integration between gamification theory and LLM system design is required to ensure that adaptive capabilities meaningfully support learning and behavioural objectives. Finally, ethics and reliability considerations must be treated as foundational design constraints rather than secondary concerns.

The following subsections further structure the discussion around the three research questions (RQ1–RQ3), providing a focused analysis of application contexts, gamification mechanics and reported benefits and challenges.

Table 6. *Summary of benefits, challenges, gamification mechanics and LLM-driven roles/functions identified across reviewed studies.*

Cluster / theme	Focus	Key benefit	Key challenge / risk	Gamification mechanics	LLM roles / functions	Representative studies
Cluster 1: Engagement & narrative immersion	Dialogue, storytelling, motivation	Enhanced immersion through contextual and emergent dialogue	Narrative drift; superficial motivation	Branching dialogue; reward-based dialogue; story-driven learning	Dynamic NPC interaction; adaptive storytelling	Sissler (2024); Li et al. (2025); Wu et al. (2025); Ronagh Nikghalb and Cheng (2025)
Cluster 2: Personalisation & adaptive support	Feedback, tutoring, learner modelling	Personalised feedback and adaptive learning support	Over-adaptation to learner behaviour	Feedback loops; adaptive feedback	Adaptive tutor; error interpretation; response tailoring	Bany Abdelnabi et al. (2025); Neshaei et al. (2025); H. Wang et al. (2025)
Cluster 3: Collaboration & social simulation	Teamwork, trust, peer interaction	Realistic collaboration and social interaction	Overtrust; anthropomorphism	Trust scores; peer-based gamification	Peer simulation; trust modelling; social delegation	Leong et al. (2024); L. Wang et al. (2025); Wester et al. (2024)
Cluster 4: Scalability & behaviour simulation	Large-scale modelling, testing, automation	Scalable user and persona simulation	Oversimplification; limited evaluation metrics	Simulation loops; roleplay scenarios	Persona generation; behaviour modelling	Khanshan et al. (2024); Spitale et al. (2025); Z. Lin et al. (2025); Crandall & Crandall (2024)
Cluster 5: Ethics, reliability & systemic risks	Trustworthiness, governance	Awareness of systemic and ethical risks	Hallucination; bias; misinformation; privacy concerns	N/A	Risk identification; governance framing	Song et al. (2025); Baldry et al. (2024)

5.7 Findings related to RQ1: Use of LLMs in gamified systems

The reviewed literature reveals that LLMs serve as intelligent agents that enhance interactivity and personalisation in gamified systems. In education, for instance, Bany Abdelnabi et al. (2025) and George et al. (2025) have demonstrated how LLMs serve as adaptive tutors, offering feedback tailored to individual learners' responses. Similarly, Wang et al. (2025) explored their use of collaborative peers to facilitate discussion and reflective thinking. In game development, LLMs such as GPT-3.5 have replaced static NPCs with responsive dialogue agents. Sissler (2024) showed how integrating LLMs into Unity enables NPCs to conduct open-ended conversations, enriching narrative engagement. Beyond gaming, LLMs are utilised for simulation and testing. Khanshan et al. (2024) introduced Prompt4Code, which models realistic user behaviour without needing actual participants, while Leong et al. (2024) implemented digital twins for virtual meetings, representing user personas through natural LLM-driven interactions.

These examples illustrate how LLMs are increasingly employed across domains. In education, they act as adaptive tutors providing personalised feedback, while in simulation contexts, they drive NPCs capable of real-time dialogue. In storytelling, researchers such as Mauri et al. (2024) have demonstrated how adaptive narratives can be generated using GPT-based models, further extending the interactive potential of gamified systems.

5.8 Findings related to RQ2: Common gamification mechanics enhanced by LLMs

The gamification strategies supported by LLMs tend to focus on adaptive, interactive and narrative-driven mechanics rather than traditional scorekeeping or badges. In social simulations, LLMs facilitate branching dialogues that allow users to make ethically nuanced decisions. Ronagh Nikghalb and Cheng (2025) demonstrated this through emergent conversations shaped by player input. Neshaei et al. (2025) applied adaptive feedback loops in learning environments, turning user mistakes into game-like opportunities for correction and learning.

Wang et al. (2025) explored how feedback derived from human interactions allows LLMs to refine gamified progression systems. Additionally, LLMs have been utilised in roleplaying and testing environments, facilitating decision-making simulations. Lin et al. (2025) and Crandall and Crandall (2024) have demonstrated how these models simulate complex social behaviour at scale.

In systems where avatars represent users, Leong et al. (2024) and Wester et al. (2024) have highlighted the inclusion of gamified metrics, such as trust scores, social delegation and collaborative achievements. Yan et al. (2025) illustrated LLM-driven storytelling, where narrative paths shift in response to user decisions, enhancing narrative gamification.

Across various environments, LLMs have been utilised to support both structural and content-based gamification. Structural mechanics such as points, badges and leaderboards have been enhanced with adaptive prompts generated by LLMs, making these features more responsive to user performance. At the same time, content-focused approaches—such as roleplaying scenarios and story-driven learning—have been expanded with LLM-driven narratives. Yan et al. (2025) illustrated this through dynamic storytelling, where narrative paths shifted in response to user decisions. Similarly, Cagnazzo et al. (2023) highlighted how content gamification integrated with LLMs allows more immersive roleplay interactions, enriching both the depth and flexibility of gamified experiences.

5.9 Findings related to RQ3: Benefits and challenges of LLM-driven gamified systems

One of the most notable benefits is the natural conversational engagement facilitated by LLMs. Li et al. (2025) and Meyerson et al. (2024) have shown that responsive dialogue enhances immersion and realism in both educational and recreational settings, enabling users to interact in ways that closely resemble human-to-human exchanges. This conversational ability not only boosts motivation but also maintains longer-term user engagement by reducing repetitive or scripted interactions.

LLMs also offer a unique advantage in scalability. For example, Lin et al. (2025) demonstrated in SimSpark that LLMs can support large-scale simulations without requiring live participants, thereby reducing resource constraints while still offering authentic interactive experiences. Scalability ensures that gamified systems can accommodate diverse learning cohorts, distributed teams or large player populations without diminishing quality.

Personalisation emerges as another recurring strength. Bany Abdelnabi et al. (2025) observed that LLMs are capable of tailoring instructional content to suit learner profiles, adapting feedback and complexity levels to individual progress. In social and collaborative environments, Leong et al. (2024) and Wester et al. (2024) have highlighted how LLM-driven agents can act as decision-making partners, effectively simulating human behaviour and fostering dynamic group interactions. This adaptability not only improves learning outcomes but also strengthens engagement in multi-user or role-based scenarios.

Innovation is further amplified by the role of LLMs in storytelling and narrative generation. By dynamically crafting branching storylines and creating NPCs with lifelike dialogue patterns, LLMs move gamified systems beyond traditional mechanics. Such innovations enable users to experience highly immersive environments where their choices and actions shape unique narrative paths, offering opportunities for creativity and deep engagement.

Despite these benefits, several challenges persist. One recurrent issue is content hallucination, the generation of inaccurate or ungrounded information that can undermine user trust and learning outcomes in evaluative or instructional contexts. Song et al. (2025) identified hallucination as a major barrier in educational applications, where factual reliability is critical. However, within gamified and narrative-driven environments, hallucination is more nuanced and can be distinguished between constructive hallucination, such as playful dialogue, creative expression or emergent storytelling, and harmful hallucination, which involves misinformation, false authority or misleading feedback. Sissler (2024) observed that prolonged interactions may lead to narrative drift and tonal inconsistency that weaken immersion; nevertheless, bounded creative variation may also enhance engagement when appropriately framed. These findings suggest that hallucination should be treated as a context-dependent design challenge rather than a uniformly negative failure mode.

Beyond technical limitations, pressing ethical concerns remain. Users may incorrectly attribute emotional intelligence or moral reasoning to LLM agents, blurring the boundaries between humans and machines. Wester et al. (2024) warned about these risks in sensitive domains such as mental health, where misplaced trust could have harmful consequences. To mitigate such issues, Ni et al. (2024) proposed using dynamic gamified assessments to evaluate the social cognition of LLMs, while Mauri et al. (2024) introduced the concept of policy sandboxing to foster empathy and inclusivity in system design.

In addition to these concerns, several studies have reported mixed or adverse outcomes when LLM-driven gamified systems were deployed without sufficient alignment between game mechanics, system goals and user expectations. In some cases, increased conversational richness did not translate into improved learning or task performance, particularly when users focused on exploiting dialogue systems for rewards rather than engaging with underlying content. Other studies have noted overreliance on adaptive feedback, where users deferred judgment to the system instead of developing independent problem-solving skills. Such findings suggest that the motivational affordances of LLMs can, under certain conditions, conflict with pedagogical or behavioural objectives. These contradictions highlight an important design tension between engagement optimisation and meaningful learning, reinforcing the need for intentional calibration of gamification mechanics, transparency of AI limitations and complementary human or rule-based oversight.

Overall, literature reflects a balance between opportunities and limitations. LLMs enhance engagement through natural dialogue, enable personalisation by adapting to user needs, provide scalability in simulation-driven contexts and support innovation in narrative design. At the same time, challenges such as hallucinations, high computational requirements, limited evaluation metrics and ethical risks underscore the importance of cautious, responsible implementation. A summary of these benefits and challenges is presented in Table 7.

Table 7. *Summary of benefits and challenges of LLM-driven gamified systems (RQ3).*

Category	Benefits	Challenges
Engagement	Natural language dialogue enhances immersion, realism and sustained user motivation by reducing scripted or repetitive interactions.	Content hallucination and narrative drift may undermine user trust, consistency and perceived credibility.
Personalisation	Adaptive, context-aware feedback tailored to individual learner profiles and progression supports improved engagement and learning outcomes.	High computational, energy and operational costs associated with model fine-tuning, inference and maintenance.
Scalability	Enables large-scale simulations and interactions without dependence on live human participants, supporting distributed and high-volume use cases.	Limited availability of robust, standardised evaluation metrics for assessing long-term learning, behavioural or engagement impact.
Innovation	Facilitates dynamic storytelling, emergent narratives and lifelike NPC behaviour, enabling creative and non-linear gamified experiences.	Ethical risks including bias propagation, misinformation, over-reliance on AI feedback and blurred human–AI boundaries.

5.10 Limitations of the review

The most obvious limitation of this review is that it may not include some useful papers. This is because only four databases were selected as sources, bringing the total to 27 papers. Besides, some papers might as well have been missed during the selection, mainly because there is an unsuitable choice of title since the screening for paper selection starts with the paper title, followed by the abstract and keywords. Next, one of the most commonly used methods for selecting papers, the forward and backward snowballing technique, was not employed because most of the selected papers were older than 5 years, i.e., older than 2018. However, these limitations can be mitigated by expanding the publication year range and adding more sources to the paper selection list.

This review is based on 27 full-text papers related to the use of LLMs in gamified systems. However, not all articles were peer-reviewed—some were preprints or open-access sources, which may affect the validity of their findings.

Furthermore, many projects are still in the prototype phase, with evaluations based on short-term studies or author-driven observations rather than longitudinal research.

Reproducibility is another issue. The majority of studies rely on proprietary LLMs, particularly OpenAI's GPT models, which hinder replication and transparency. Ultimately, while ethical concerns are often acknowledged, practical solutions are often limited. Most studies lack concrete frameworks for managing bias, misinformation or user overtrust, indicating an underdeveloped area in ethical design.

6 CONCLUSION

This systematic literature review analysed 27 scholarly articles on the integration of large language models (LLMs) with gamified systems, offering critical insights into their transformative impact on engagement and interactivity. The review bridges two previously distinct research domains by integrating LLMs with gamification and demonstrates how their convergence enables adaptive, personalised and human-like interactions that extend well beyond traditional point-and-badge mechanics. By synthesising findings across education, simulation and storytelling contexts, this review highlights the capacity of LLMs to redefine user engagement through dynamic narratives, real-time feedback and intelligent behaviour modelling.

Regarding the guiding research questions, three major themes emerged. Firstly, LLMs contribute significantly to user modelling, adaptive dialogue and personalised feedback (RQ1), positioning them as intelligent agents within gamified systems. Secondly, their integration reconfigures gamification mechanics (RQ2), shifting from static reward structures to richer elements, such as branching narratives, collaborative roleplay and error-based feedback loops. Thirdly, the literature identifies both benefits and challenges (RQ3). Benefits include enhanced engagement, scalability in simulation environments and the ability to simulate human-like decision-making. Conversely, recurring challenges include content hallucinations, ethical concerns surrounding bias and over-reliance and the lack of design frameworks to ensure alignment with user needs.

Despite these contributions, the review identifies notable research gaps. Existing studies tend to emphasise proof-of-concept or short-term implementations, with limited attention to longitudinal evaluation, standardised assessment frameworks and robust ethical safeguards. Furthermore, while the adaptability of LLMs holds great promise, their effective deployment requires attention to computational costs, transparency and user trust.

Future research should therefore prioritise ethical and user-centred design approaches, integrate risk mitigation strategies and adopt rigorous evaluation methods to measure long-term effectiveness. By addressing these gaps, the research community can ensure that systems developed through the integration of LLMs with gamification not only advance interactivity and personalisation but also remain sustainable, equitable and responsible in practice.

ADDITIONAL INFORMATION AND DECLARATIONS

Acknowledgments: The authors appreciate the reviewers' thoughtful and rigorous comments, which enhanced the manuscript. The authors would also like to thank everyone who contributed, especially the *Faculty of Computer Science and Information Technology* at *Universiti Malaysia Sarawak* (UNIMAS) for this research opportunity.

Funding: This work was funded by the UNIMAS.

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: N.J.: Data curation, Methodology, Writing – original draft, Writing – review and editing. L.G.L.: Data curation, Writing – original draft. C.W.S.: Writing – review and editing. S.B.Y.: Methodology, Writing – review and editing. S.N.J.: Conceptualization, Methodology.

Statement on the Use of Artificial Intelligence Tools: The authors acknowledge the use of generative AI tools in the preparation of this manuscript. Specifically, AI was used to assist with synthesising literature review findings, structuring comparative tables, and generating thematic content for the Discussion section. Additionally, generative AI tools were used to create the conceptual illustrations and thematic clusters presented in Figures 3 and 4. The authors have verified the accuracy, validity, and appropriateness of all AI-generated content and citations. Human authors remain fully responsible for the entire content of the manuscript.

Data Availability: The data that support the findings of this study are available from the corresponding author.

APPENDIX A – PRISMA 2020 CHECKLIST

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Section 1
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Sections 3.1–3.1.1
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Section 3.1.5
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Sections 3.1.3 and 3.2
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Section 3.1.4
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Section 3.2.1
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Section 3.2.2
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Section 3.2.2
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Section 3.2.2
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	See important note below
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Not applicable (qualitative synthesis)
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Sections 3.2.1–3.2.2
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Section 3.2.2
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Sections 3.2.3 and 4; Figure 1
	13d	Describe any methods used to synthesise results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Sections 3.2.2 and 4.1–4.3
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Not applicable
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesised results.	Not Reported
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Not Reported

Section and Topic	Item #	Checklist item	Location where item is reported
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Not Reported
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Section 3.2.1; Figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Section 3.2.1
Study characteristics	17	Cite each included study and present its characteristics.	Sections 4.1–4.3; Tables 3–5; Appendix B
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	See important note below
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Sections 4.1–4.3
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Sections 4.1–4.3 and 5.1–5.5
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Not applicable (qualitative synthesis)
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Not applicable
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesised results.	
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Not reported
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Not reported
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Section 5
	23b	Discuss any limitations of the evidence included in the review.	Section 5.10
	23c	Discuss any limitations of the review processes used.	Section 5.10
	23d	Discuss implications of the results for practice, policy, and future research.	Sections 5.6 and 6
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Not registered
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Not reported
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Not applicable
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Additional Information and Declarations – Funding
Competing interests	26	Declare any competing interests of review authors.	Additional Information and Declarations – Conflict of Interests
Availability of data, code	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Additional Information and

Section and Topic	Item #	Checklist item	Location where item is reported
and other materials			Declarations – Data Availability

APPENDIX B – LIST OF ALL INCLUDED PAPERS (N = 27)

Authors	Year	Title
Ataguba, G., Oyebode, O., Chan, G., & Orji, R.	2025	MyShoppingBuddy: Exploring Persuasive Strategies and Design Opportunities for Gamifying Real World Shopping Experiences
Baldry, M. K., Happa, J., Steed, A., Smith, S., & Glencross, M.	2024	From Embodied Abuse to Mass Disruption: Generative, Inter-Reality Threats in Social, Mixed-Reality Platforms
Bany Abdelnabi, A. A., Soykan, B., Bhatti, D., & Rabadi, G.	2025	Usefulness of Large Language Models (LLMs) for Student Feedback on H&P During Clerkship: Artificial Intelligence for Personalized Learning
Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D.	2020	Language Models Are Few-Shot Learners
Cagnazzo, C., Garaccione, G., Coppola, R., Ardito, L., & Torchiano, M.	2023	UMLegend: A Gamified Learning Tool for Conceptual Modeling with UML Class Diagrams
Crandall, J. L., & Crandall, A. S.	2024	Large Language Model-Supported Software Testing with the CS Matrix Taxonomy
George, A., Storey, V. C., & Hong, S.	2025	Unraveling the Impact of ChatGPT as a Knowledge Anchor in Business Education
Khanshan, A., Van Gorp, P., & Markopoulos, P.	2024	Evaluation of Code Generation for Simulating Participant Behavior in Experience Sampling Method by Iterative In-Context Learning of a Large Language Model
Kumar, H., Musabirov, I., Reza, M., Shi, J., Wang, X., Williams, J. J., Kuzminykh, A., & Liut, M.	2024	Guiding Students in Using LLMs in Supported Learning Environments: Effects on Interaction Dynamics, Learner Performance, Confidence, and Trust
Leong, J., Tang, J., Cutrell, E., Junuzovic, S., Baribault, G. P., & Inkpen, K.	2024	Dittos: Personalised, Embodied Agents That Participate in Meetings When You Are Unavailable
Leslie, L. G., Jali, N., Yahia, S. Ben, Wai Shiang, C., & Junaini, S. N.	2025	Exploring Large Language Models in Gamified Environments: A Systematic Literature Review
Li, Y., Sun, L., & Zhang, Y.	2025	MetaAgents: Large Language Model Based Agents for Decision-Making on Teaming
Lin, J., Dai, X., Xi, Y., Liu, W., Chen, B., Zhang, H., Liu, Y., Wu, C., Li, X., Zhu, C., Guo, H., Yu, Y., Tang, R., & Zhang, W.	2025	How Can Recommender Systems Benefit from Large Language Models: A Survey
Lin, Z., Shan, Y., Gao, L., Jia, X., & Chen, S.	2025	SimSpark: Interactive Simulation of Social Media Behaviors
Mauri, A., Hsu, Y. C., Verma, H., Tocchetti, A., Brambilla, M., & Bozzon, A.	2024	Policy Sandboxing: Empathy As An Enabler Towards Inclusive Policy-Making
Meyerson, E., Nelson, M. J., Bradley, H., Gaier, A., Moradi, A., Hoover, A. K., & Lehman, J.	2024	Language Model Crossover: Variation through Few-Shot Prompting
Neshaei, S. P., Tolzin, A., Berkle, Y., Leuchter, M., Leimeister, J. M., Janson, A., & Wambsganss, T.	2025	Leveraging Learner Errors in Digital Argumentation Learning: How ALure Helps Students Learn from their Mistakes and Write Better Arguments
Ni, Q., Yu, Y., Ma, Y., Lin, X., Deng, C., Wei, T., & Xuan, M.	2024	The Social Cognition Ability Evaluation of LLMs: A Dynamic Gamified Assessment and Hierarchical Social Learning Measurement Approach
Ronagh Nikghalb, M., & Cheng, J.	2025	Interrogating AI: Characterizing Emergent Playful Interactions with ChatGPT
Sissler, J.	2024	Enhancing Non-Player Characters in Unity 3D Using GPT-3.5
Song, T., Jamieson, J., Zhu, T., Yamashita, N., & Lee, Y. C.	2025	From Interaction to Attitude: Exploring the Impact of Human-AI Cooperation on Mental Illness Stigma

Authors	Year	Title
Spitale, M., Axelsson, M., & Gunes, H.	2025	VITA: A Multi-Modal LLM-Based System for Longitudinal, Autonomous and Adaptive Robotic Mental Well-Being Coaching
Wang, H., Wu, C., Huang, Y., & Qi, T.	2025	Learning Human Feedback from Large Language Models for Content Quality-aware Recommendation
Wang, L., Zhang, J., Yang, H., Chen, Z. Y., Tang, J., Zhang, Z., Chen, X., Lin, Y., Sun, H., Song, R., Zhao, X., Xu, J., Dou, Z., Wang, J., & Wen, J. R.	2025	User Behavior Simulation with Large Language Model-based Agents
Wester, J., Pohl, H., Hosio, S., & van Berkel, N.	2024	"This Chatbot Would Never...": Perceived Moral Agency of Mental Health Chatbots
Wu, T., Zhu, J., Zhou, W., & Li, H.	2025	RESIST: Rationale-Enhanced and Reward Model-Based End-to-End Social Influence Dialogue System
Yan, Z., & Xiang, Y.	2025	Social Life Simulation for Non-Cognitive Skills Learning

REFERENCES

Ataguba, G., Oyebode, O., Chan, G., & Orji, R. (2025). MyShoppingBuddy: Exploring Persuasive Strategies and Design Opportunities for Gamifying Real World Shopping Experiences. *ACM Games: Research and Practice*, 3(3), 1–34. https://doi.org/10.1145/3735141

Baldry, M. K., Happa, J., Steed, A., Smith, S., & Glencross, M. (2024). From Embodied Abuse to Mass Disruption: Generative, Inter-Reality Threats in Social, Mixed-Reality Platforms. *Digital Threats: Research and Practice*, 5(4), 1–36. https://doi.org/10.1145/3696015

Bany Abdelnabi, A. A., Soykan, B., Bhatti, D., & Rabadi, G. (2025). Usefulness of Large Language Models (LLMs) for Student Feedback on H&P During Clerkship: Artificial Intelligence for Personalized Learning. *ACM Transactions on Computing for Healthcare*, 7(3), 1–19. https://doi.org/10.1145/3712298

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. In *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*. NIPS.

Cagnazzo, C., Garaccione, G., Coppola, R., Ardito, L., & Torchiano, M. (2023). UMLegend: A Gamified Learning Tool for Conceptual Modeling with UML Class Diagrams. In *Gamify 2023: Proceedings of the 2nd International Workshop on Gamification in Software Development, Verification, and Validation*, (pp. 2–5). ACM. https://doi.org/10.1145/3617553.3617883

Crandall, J. L., & Crandall, A. S. (2024). Large Language Model-Supported Software Testing with the CS Matrix Taxonomy. *Journal of Computing Sciences in Colleges*, 40(1), 49–58.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, (pp. 9–15). ACM. https://doi.org/10.1145/2181037.2181040

George, A., Storey, V. C., & Hong, S. (2025). Unraveling the Impact of ChatGPT as a Knowledge Anchor in Business Education. *ACM Transactions on Management Information Systems*, 16(1), 1–30. https://doi.org/10.1145/3705734

Khanshan, A., Van Gorp, P., & Markopoulos, P. (2024). Evaluation of Code Generation for Simulating Participant Behavior in Experience Sampling Method by Iterative In-Context Learning of a Large Language Model. *Proceedings of the ACM on Human-Computer Interaction*, 8(EICS), 1–19. https://doi.org/10.1145/3661143

Kumar, H., Musabirov, I., Reza, M., Shi, J., Wang, X., Williams, J. J., Kuzminykh, A., & Liut, M. (2024). Guiding Students in Using LLMs in Supported Learning Environments: Effects on Interaction Dynamics, Learner Performance, Confidence, and Trust. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2), 1–30. https://doi.org/10.1145/3687038

Leong, J., Tang, J., Cutrell, E., Junuzovic, S., Baribault, G. P., & Inkpen, K. (2024). Dittos: Personalised, Embodied Agents That Participate in Meetings When You Are Unavailable. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2), 1–28. https://doi.org/10.1145/3687033

Leslie, L. G., Jali, N., Yahia, S. Ben, Wai Shiang, C., & Junaini, S. N. (2025). Exploring Large Language Models in Gamified Environments: A Systematic Literature Review. In *2025 14th International Conference on Information Technology in Asia* (pp. 119–124). IEEE. https://doi.org/10.1109/CITA66455.2025.11198821

Li, Y., Sun, L., & Zhang, Y. (2025). MetaAgents: Large Language Model Based Agents for Decision-Making on Teaming. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), 1–27. https://doi.org/10.1145/3711032

Lin, J., Dai, X., Xi, Y., Liu, W., Chen, B., Zhang, H., Liu, Y., Wu, C., Li, X., Zhu, C., Guo, H., Yu, Y., Tang, R., & Zhang, W. (2025). How Can Recommender Systems Benefit from Large Language Models: A Survey. *ACM Transactions on Information Systems*, 43(2), 1–47. https://doi.org/10.1145/3678004

Lin, Z., Shan, Y., Gao, L., Jia, X., & Chen, S. (2025). SimSpark: Interactive Simulation of Social Media Behaviors. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), 1–32. https://doi.org/10.1145/3711066

Mauri, A., Hsu, Y. C., Verma, H., Tocchetti, A., Brambilla, M., & Bozzon, A. (2024). Policy Sandboxing: Empathy As An Enabler Towards Inclusive Policy-Making. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2), Article 369. https://doi.org/10.1145/3686908

Meyerson, E., Nelson, M. J., Bradley, H., Gaier, A., Moradi, A., Hoover, A. K., & Lehman, J. (2024). Language Model Crossover: Variation through Few-Shot Prompting. *ACM Transactions on Evolutionary Learning and Optimization*, 4(4). https://doi.org/10.1145/3694791

Neshaei, S. P., Tolzin, A., Berkle, Y., Leuchter, M., Leimeister, J. M., Janson, A., & Wambsganss, T. (2025). Leveraging Learner Errors in Digital Argumentation Learning: How ALure Helps Students Learn from their Mistakes and Write Better Arguments. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), 1–32. https://doi.org/10.1145/3711023

Ni, Q., Yu, Y., Ma, Y., Lin, X., Deng, C., Wei, T., & Xuan, M. (2024). The Social Cognition Ability Evaluation of LLMs: A Dynamic Gamified Assessment and Hierarchical Social Learning Measurement Approach. *ACM Transactions on Intelligent Systems and Technology*, 16(6), Article 135. https://doi.org/10.1145/3673238

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372, n160. https://doi.org/10.1136/bmj.n160

Ronagh Nikghalb, M., & Cheng, J. (2025). Interrogating AI: Characterizing Emergent Playful Interactions with ChatGPT. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), 1–23. https://doi.org/10.1145/3711015

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human Computer Studies*, 74, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006

Sissler, J. (2024). Enhancing Non-Player Characters in Unity 3D using GPT-3.5. *ACM Games: Research and Practice*, 2(3), 1–16. https://doi.org/10.1145/3662003

Song, T., Jamieson, J., Zhu, T., Yamashita, N., & Lee, Y. C. (2025). From Interaction to Attitude: Exploring the Impact of Human-AI Cooperation on Mental Illness Stigma. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), 1–31. https://doi.org/10.1145/3710987

Spitale, M., Axelsson, M., & Gunes, H. (2025). VITA: A Multi-Modal LLM-Based System for Longitudinal, Autonomous and Adaptive Robotic Mental Well-Being Coaching. *ACM Transactions on Human-Robot Interaction*, 14(2), 1–28. https://doi.org/10.1145/3712265

Wang, H., Wu, C., Huang, Y., & Qi, T. (2025). Learning Human Feedback from Large Language Models for Content Quality-aware Recommendation. *ACM Transactions on Information Systems*, 43(4), 1–28. https://doi.org/10.1145/3727144

Wang, L., Zhang, J., Yang, H., Chen, Z. Y., Tang, J., Zhang, Z., Chen, X., Lin, Y., Sun, H., Song, R., Zhao, X., Xu, J., Dou, Z., Wang, J., & Wen, J. R. (2025). User Behavior Simulation with Large Language Model-based Agents. *ACM Transactions on Information Systems*, 43(2), 1–37. https://doi.org/10.1145/3708985

Wester, J., Pohl, H., Hosio, S., & van Berkel, N. (2024). “This Chatbot Would Never...”: Perceived Moral Agency of Mental Health Chatbots. *Proceedings of the ACM on Human-Computer Interaction*, 8(1), 1–28. https://doi.org/10.1145/3637410

Wu, T., Zhu, J., Zhou, W., & Li, H. (2025). RESIST: Rationale-Enhanced and Reward Model-Based End-to-End Social Influence Dialogue System. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 21(6), 1–23. https://doi.org/10.1145/3736580

Yan, Z., & Xiang, Y. (2025). Social Life Simulation for Non-Cognitive Skills Learning. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), 1–44. https://doi.org/10.1145/3711068

Acta Informatica Pragensia is published by the Prague University of Economics and Business, Czech Republic | eISSN: 1805-4951