INTERACTIVE LEARNING ENVIRONMENTS
https://doi.org/10.1080/10494820.2026.2658204

Routledge
Taylor & Francis Group

Human–AI interaction in a socio-educational metaverse: insights from a developmental evaluation of AI avatars

Manuel B. Garcia [a,b,c]

[a]College of Education, University of the Philippines Diliman, Quezon City, Philippines; [b]Educational Innovation and Technology Hub, FEU Institute of Technology, Manila, Philippines; [c]Graduate School of Education, Korea University, Seoul, South Korea

ABSTRACT

The metaverse and artificial intelligence (AI) are increasingly intersecting in educational contexts, yet limited empirical research has examined how generative AI avatars function within socially interactive virtual environments. This study investigates the deployment of generative AI avatars within a socio-educational metaverse environment. Using a developmental evaluation approach, data were collected through interviews with seven institutional stakeholders, teacher-generated reflections, internal documentation, embedded user feedback captured through in-platform reporting tools, and longitudinal field memos across an iterative deployment cycle. Findings indicate that the transition from scripted NPCs to generative AI avatars recalibrated users' attribution of agency, intensified dialogic unpredictability, and elevated social realism beyond visual fidelity. Voice-mediated interaction emerged as a threshold mechanism for co-presence, while algorithmic improvisation exposed tensions between pedagogical intent and stochastic response generation. The deployment further revealed affective frictions, expectation misalignments, and the mediating role of AI literacy in shaping trust, participation, and interpretive coherence. Overall, the study advances a sociotechnical understanding of AI avatars as co-constructors of meaning and interaction, offering implications for the design, implementation, and governance of future AI-enhanced metaverse learning environments.

ARTICLE HISTORY
Received 5 December 2025
Accepted 2 April 2026

KEYWORDS
Metaverse; learning environment; artificial intelligence; generative AI; AI in education; AI avatars

Introduction

Artificial intelligence (AI) has evolved from expert systems and rule-based logic into a diverse suite of computational agents that now permeate educational ecosystems. Early forms of AI in education focused on intelligent tutoring systems, learner modeling, and data-driven feedback, which enabled automation in tasks such as assessment and adaptive content delivery (Castillo-Martínez et al., 2024; Wang et al., 2024). Recent advances in machine learning, particularly through transformer-based architectures like GPT, have ushered in a generative phase that enables the analysis and production of meaningful output across modalities. Generative AI (GenAI) now underpins a range of instructional applications including personalized tutoring, formative feedback, dialogic support, and content generation (Bozkurt et al., 2024; Labadze et al., 2023; Xiaoyu et al., 2025). Systematic reviews reported that these affordances have been shown to foster engagement, creativity, and critical thinking while simultaneously raising concerns over data bias, ethical governance, and pedagogical transparency (Belkina et al., 2025; Izquierdo-Álvarez & Jimeno-Postigo, 2025). Concerns over factual inaccuracies (Acut et al., 2025), bias amplification (Xiao et al., 2025), and the erosion of learner agency (Miranda et al., 2025) have also prompted calls for deeper scrutiny of GenAI's role in shaping educational experience and pedagogical intent. As GenAI increasingly occupies the interstitial space between learner cognition and instructional design, there is a critical need to examine how it reconfigures notions of agency, intentionality, and authorship in teaching and learning contexts.

Parallel to the rise of generative AI, the metaverse has been increasingly positioned as an immersive learning ecosystem distinguished by spatial continuity, embodied interaction, and simulated co-presence (Onu et al., 2024). Its applications in education have expanded rapidly, with institutions leveraging virtual campuses,

CONTACT Manuel B. Garcia manuelgarciaph@yahoo.com; mbgarcia@feutech.edu.ph

collaborative simulations, and gamified tasks to enhance learner engagement and support experiential pedagogies (Tlili et al., 2022). Within these environments, interactive agents have often served as critical infrastructure for guiding users, structuring activities, and scaffolding participation. However, early implementations largely relied on non-playable characters (NPCs) operating through pre-scripted dialogue trees and rigid behavioral routines (Almeman et al., 2025). These artificial agents served primarily as environmental scaffolds, offering instructional cues or procedural guidance within tightly controlled virtual tasks. While effective for low-variability scenarios such as procedural simulations or basic orientation modules, their inability to sustain adaptive and responsive interaction limited their epistemic utility in more dialogic or exploratory pedagogies. This constraint has been identified as particularly salient in metaverse-based educational environments where relational dynamics, improvisation, and social presence are central to learner engagement and identity formation (Garcia, 2025). As educational institutions increasingly adopt metaverse platforms for collaborative learning, the integration of AI into avatar behavior introduces new possibilities for enhancing interactivity and presence (Adarkwah et al., 2024), prompting a need to reassess long-held assumptions about scripted agents and their pedagogical role.

Despite accelerating developments in both generative AI and immersive education, a critical research gap persists in understanding the deployment of conversational AI agents within socially situated metaverse platforms. Existing literature has largely bifurcated its focus between visual realism of avatars and domain-specific intelligent agents, often overlooking the systemic ramifications of embedding large language models into persistent virtual environments (Al-Emran, 2024). Recent reviews have outlined the potential of generative AI to serve as virtual tutors, teaching assistants, and peer collaborators in the metaverse, yet these conceptual roles remain underexplored in empirical settings where unpredictability, social dynamics, and learner improvisation prevail (Ansari et al., 2024; Koohang et al., 2023). Moreover, studies have flagged ethical, pedagogical, and technical concerns (e.g. information reliability, digital literacy asymmetries, and diminished human empathy) that become amplified in fully immersive simulations (Bozkurt et al., 2024; Lv, 2023). This study responds to these limitations by examining what is learned when NPCs are supplanted by generative AI avatars capable of maintaining voice-based and context-aware interactions within open-ended virtual spaces. Specifically, it seeks to unpack the sociotechnical frictions and affordances that arise when scripted behavioral scaffolds give way to generative improvisation. Illuminating how AI avatars recalibrate the pedagogical terrain of metaverse learning environments is essential for advancing theoretical models of AI-mediated education and informing the design of future immersive systems that prioritize relational authenticity, learner agency, and contextual adaptability.

Materials and methods

Research context

This study was conducted within the framework of a social-oriented metaverse environment that underwent a transition from NPCs to AI-driven avatars. MILES Virtual World is a metaverse-based platform originally developed to simulate campus environments and foster student engagement through interactive digital spaces (Garcia et al., 2023a). Initially limited to static campus scenarios populated by scripted NPCs, the platform gradually expanded its design scope and interactive affordances. Over successive iterations, it evolved into a broader virtual ecosystem intended to support socially dynamic experiences through the integration of generative AI. The current implementation also marked the first deployment beyond the institutional campus through the introduction of a new "summer camp" environment (see Figure 1), which is conceptualized as an extended map within the virtual world (Garcia et al., 2024). This off-campus setting featured a range of playful and socially oriented affordances including kayaking, fishing, competitive racing, and water-based mini-games (see Figure 2 for an example activity in the virtual world). These playful and unstructured contexts were intentionally selected to examine how generative AI avatars function in socially emergent rather than tightly scripted instructional settings. The environment allowed observation of dialogic improvisation, informal learning exchanges, and affective engagement beyond formal curricular constraints. Overall, the setting served as a naturalistic testbed for examining human–AI interaction in socially situated learning contexts rather than as a structured instructional intervention.

Figure 1. MILES virtual world summer camp environment.

System innovation and design motivation

In transitioning from NPCs to generative AI-powered entities, the project aimed to surface implementation-level learning, operational heuristics, and systemic enablers that informed both current and future design trajectories. This transformation was intended to shift the role of virtual agents from scripted presences to quasi-autonomous avatars capable of simulating naturalistic human interaction. The strategic pivot toward AI integration was also catalyzed by the functional limitations of prior NPC configurations observed in earlier versions of the MILES Virtual World (Garcia et al., 2023b). While these scripted agents facilitated rudimentary interaction loops, they lacked dialogic plasticity and contextual responsiveness. The improvements leveraged transformer-based natural language models via API deployment. Specifically, the system utilized OpenAI's GPT-4 architecture (GPT-4o), configured with controlled temperature settings to balance coherence and variability. A structured system-level prompt was used to define role boundaries, tone expectations, and topical scope. No fine-tuning was conducted and behavior was shaped through prompt engineering and interaction constraints. Average response latency ranged between approximately 1.5–3 s under stable network conditions, remaining within conversational tolerance for voice interaction. Text-to-speech and speech-to-text modalities were integrated to enable voice-based exchanges. This architectural augmentation not only addressed prior design deficiencies but also introduced new vectors for social realism,

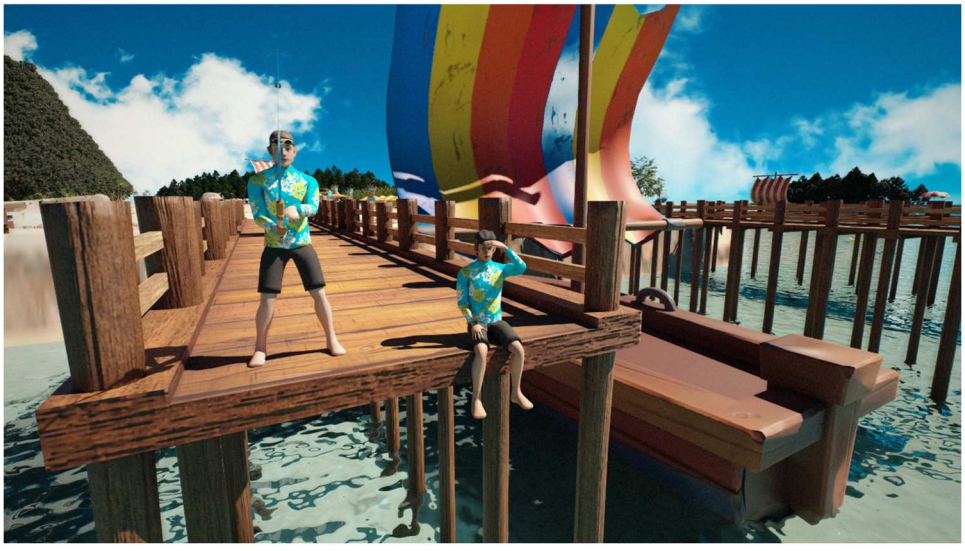

Figure 2. Users engaging in a fishing simulation within the socio-educational metaverse.

engagement, and emergent user behavior. From a systems perspective, this innovation operated at the intersection of human–computer interaction, game-based learning, and AI-mediated communication.

Developmental evaluation approach

Building on the system innovations and iterative design shifts described in the preceding section, this study adopted a Developmental Evaluation (DE) approach. DE is a methodology specifically designed to support the evolution of innovations in complex, uncertain, and dynamic environments (Patton, 2011). Unlike formative or summative approaches, DE embeds evaluative thinking into the design and implementation process. This approach makes it particularly suitable for projects like AI-enhanced metaverse platforms, where technical architecture, user behaviors, and design goals evolve simultaneously (Almeman et al., 2025; Lv, 2023). The use of DE in this study was also driven by several contextual contingencies, such as the novelty of generative AI integration, the fluidity of user interaction patterns, and the iterative nature of platform development. Rather than relying on retrospective assessments, DE facilitated a continuous feedback loop between stakeholder insights and system modifications. To structure the developmental process, the evaluation followed five interconnected phases (see Figure 3). These phases provided a flexible scaffold that aligned evaluative inquiry with the iterative nature of the innovation and allowed for the documentation of design rationales, identification of inflection points, and evaluation of both anticipated and emergent outcomes.

Data collection and analysis

Data collection was conducted longitudinally and iteratively, in alignment with DE principles emphasizing adaptive responsiveness and embedded inquiry. Semi-structured interviews were administered with key stakeholders, including two technical professionals specializing in game development and 3D modeling, three instructional faculty serving as pedagogical interlocutors, and two administrators overseeing institutional digital transformation. These interviews explored design rationales, affordance negotiation, and emergent tensions (Miles et al., 2013). Supplementary data were derived from teacher-generated reflections based on exploratory interactions with the environment, internal documentation (e.g. changelogs, version histories, design schematics), and user-contributed feedback obtained via embedded reporting tools. Additionally, reflexive field memos authored by the lead investigator captured critical incidents, project inflection points, and researcher positionality throughout the design cycle (Ortlipp, 2008). Following DE principles of sensemaking over static coding, collected data were analyzed using a retrospective real-time mapping technique. An inductive thematic analysis (Braun & Clarke, 2006) was conducted across all data sources to identify recurring patterns and emergent insights. Themes were organized according to a timeline of development milestones, allowing the research team to trace how specific design decisions influenced later outcomes. Where applicable, findings were also interpreted through the lens of complexity-aware evaluation, which emphasizes understanding nonlinear pathways, unintended consequences, and contextual dependencies (Rogers, 2008). The iterative engagement with developers, teachers, and institutional stakeholders enabled a broader understanding of how implementation decisions shaped not only user experience but institutional attitudes toward AI integration.

Figure 3. Five phases of the developmental evaluation framework.

Ethical considerations

Participation in interviews and teacher reflections was voluntary, and informed consent was obtained from all interview participants prior to data collection. Participation or non-participation had no impact on academic standing, institutional roles, or access to the metaverse platform. Users interacting with the metaverse environment were informed that they were engaging with AI-generated avatars rather than human agents, and no deceptive framing regarding avatar identity or capabilities was employed. The deployment did not involve experimental manipulation, behavioral intervention, or psychological deception. Data derived from interviews, reflections, embedded feedback tools, and documentation were anonymized prior to analysis. No personally identifiable information is reported in this manuscript. The study complied with institutional guidelines for responsible research conduct and ethical data handling.

Results

Descriptive interaction metrics

System logs were analyzed to quantify observable shifts following AI avatar integration (Table 1). A total of 1,284 sessions were recorded during the AI deployment phase, compared to 742 sessions during the NPC phase. Mean conversational turns per session increased from 3.4 (NPC phase) to 11.2 (AI phase). Average session duration increased from 4.8 min to 13.6 min. User-initiated follow-up prompts increased from 22% to 61% of total sessions. Voice-enabled sessions (n = 412) averaged 15.1 conversational turns compared to 9.4 turns in text-only sessions. Episodes of instructional divergence occurred in 14% of structured task sessions. Anthropomorphic attribution language appeared in approximately 18% of AI sessions. Emotional tone mismatch episodes were identified in 9% of sessions involving affective disclosure.

Empirically derived lessons from AI avatar deployment

Longitudinal deployment of generative AI avatars within the socio-educational metaverse platform produced ten recurring empirical patterns across interactional, pedagogical, technical, and institutional domains. These lessons were derived from iterative analysis of 1,284 interaction sessions, 7 stakeholder interviews, 23 reflexive field memos, embedded user feedback reports, and version-tracked system modifications. The lessons represent patterned observations rather than isolated incidents and reflect consistent phenomena observed across deployment phases. Table 2 presents the lessons identified through triangulated analysis across these data sources (see Appendix A for the cross-source contribution matrix). Illustrative participant feedback excerpts and corresponding observed impacts for each lesson are provided in Appendix B.

Complexity-aware system dynamics

The lessons identified in the study did not emerge linearly. Rather, they interacted dynamically across deployment cycles, producing feedback loops, adaptive refinements, and residual tensions characteristic of complex sociotechnical systems. For example, open-ended prompts triggered conversational branching in 47% of exploratory sessions. Prompt boundary cues introduced in Version 1.2 reduced instructional drift in

Table 1. Descriptive interaction indicators across deployment phases.

Indicator	NPC Phase	AI Avatar Phase
Total Sessions	742	1,284
Mean Turns per Session	3.4	11.2
Mean Session Duration (minutes)	4.8	13.6
User-Initiated Follow-Up Rate	22%	61%
Voice Sessions	–	412
Instructional Divergence Episodes	3%	14%
Anthropomorphic Attribution Incidents	2%	18%
Affective Mismatch Episodes	–	9%

Table 2. Empirically derived lessons from AI avatar deployment.

Lesson	Theme	Empirical Result	Primary Data Source
1	Reconstituting Agency	Users attributed responsiveness and intentionality to AI avatars during sustained interactions.	System logs, interviews
2	Emergent Dialogue	Conversations unfolded in open-ended, unpredictable ways not observed in NPC-based exchanges.	Interaction transcripts
3	Social Realism	Perceived social presence increased when responses were contextually contingent and temporally immediate.	User feedback
4	Pedagogical Collision	AI-generated responses occasionally diverged from predefined instructional pathways.	Teacher reflections
5	Voice Mediation	Voice-enabled interactions resulted in longer and more immersive engagement episodes than text-only exchanges.	System logs, user feedback
6	Affective Frictions	Mismatches in emotional tone reduced users' perceived trust and engagement.	User reports
7	Informal Learning	Playful and unstructured exchanges led to extended exploratory dialogue.	Observational memos
8	Expectation Misalignment	Users overestimated AI's instructional and emotional capacities.	Interviews
9	Ethics of Simulation	Realistic conversational behavior raised concerns about artificial intimacy and relational ambiguity.	Stakeholder reflections
10	AI Literacy Scaffolding	Limited AI literacy constrained the quality and depth of user interaction.	Interaction transcripts

structured tasks from 19% to 11%, while exploratory branching remained high in informal contexts. Voice activation increased immersion-related feedback by 38% but also amplified emotional expectation attribution. Twelve prompt refinements and four modality adjustments were documented across deployment phases. Table 3 maps the ten lessons within a complexity-aware framework.

Discussion

Within this specific socio-educational deployment, the introduction of AI avatars in an educational metaverse revealed a series of conceptual, technical, and experiential shifts that were not fully anticipated during initial design. As the system transitioned from scripted NPCs to generative AI agents, emergent forms of interaction began to challenge existing assumptions about presence, control, and user engagement in virtual environments. These insights, while situated within the MILES Virtual World implementation, are significant for understanding how AI alters interface dynamics and for informing the future development of socially responsive systems that blur the boundaries between automation and authenticity. Building on the empirical patterns reported in the Results, this section interprets the ten developmentally derived lessons through a layered sociotechnical lens, examining how technical architecture, interactional dynamics, pedagogical processes, and institutional governance intersect and co-evolve in AI-mediated environments. Rather than treating the lessons as isolated phenomena, the discussion analyzes their interdependencies, theoretical implications, and broader relevance for designing and governing generative AI systems in socio-educational metaverse contexts.

Table 3. Complexity-aware emergence matrix of AI avatar deployment.

Lesson	Trigger Condition	Emergent Pattern	Feedback Loop Activated	Adaptive Modification	Residual Tension
1	Sustained dialogue	Attribution of intentionality	Monitoring of agency language	None	Illusion vs autonomy
2	Open-ended prompts	Conversational branching	Prompt refinement	Boundary cues added	Freedom vs control
3	Context-sensitive replies	Heightened social presence	Latency monitoring	Speech tuning	Responsiveness dependency
4	Structured tasks	Instructional drift	Instructional review	Prompt scaffolding	Persistent unpredictability
5	Voice activation	Immersion amplification	Voice latency testing	Tone calibration	Empathy over-attribution
6	Emotional disclosure	Trust fluctuation	Tone review	Affective guardrails	Authenticity limits
7	Playful context	Extended exploratory dialogue	Observational monitoring	None	Assessment ambiguity
8	High reliance queries	Capability overestimation	Transparency messaging	Disclaimer embedding	Persistent overreliance
9	Intimate tone	Relational ambiguity	Ethics consultation	Disclosure reinforcement	Simulation realism tension
10	User confusion	Shallow prompting	Literacy assessment	Embedded AI cues	Uneven participation

Lesson 1: reconstituting agency through embodied AI interaction

Agency is typically defined as the perceived ability of an entity to act autonomously, make decisions, and exert influence within an environment (Garcia et al., 2023a). In transitioning from scripted NPCs to generative AI avatars, the deployment introduced a recalibration of agency wherein users began interpreting these avatars not as deterministic interfaces but as dialogic agents capable of exhibiting intentionality, responsiveness, and semantic fluidity. This perceptual shift rendered interaction less about interface control and more about relational emergence, as users encountered avatars that produced contingent responses shaped by probabilistic language models. In this lesson, agency is treated based on how users cognitively ascribe intentionality and decision-making capacity to the avatar. This analytic focus differs from later discussions of realism or presence, which concern perceptual plausibility and experiential immersion rather than perceived authorship of action. Rather than attributing agency to intrinsic properties of the AI system, this study draws from the theory of sociomateriality, which frames agency as a relational effect enacted through the entanglement of human and nonhuman actors (Suchman, 2007). The avatars' apparent intentionality emerged not from autonomous cognition but from their embeddedness within discursively rich contexts where learners sought meaning, coherence, and responsiveness. This aspect aligns with recent work framing AI agents as epistemic actors whose influence unfolds through situated interaction rather than predefined logic (Zhang & Siau, 2024). Reconstituting agency in this way challenges long-standing assumptions in instructional design and human–computer interaction, which foregrounds the need to reconsider how intentionality and authorship are distributed in AI-mediated learning ecologies (Hwang & Won, 2022).

Lesson 2: simulated intelligence as a catalyst for emergent dialogue

Historically, conversational affordances in educational metaverses were constrained by deterministic dialogue trees that mapped user prompts to fixed textual outputs. The integration of generative AI into avatar behavior instantiated a discursive ecology in which user initiated open-ended linguistic sequences characterized by semantic unpredictability and contextual elasticity (Córdova-Esparza, 2025). This transition restructured the communicative logic of the environment, which enabled avatars to produce responses that were topically coherent, affectively modulated, and syntactically varied. Contrary to conventional agent frameworks that prioritize content stability and pedagogical containment, the AI avatars in this study operated within a spectrum of interpretive ambiguity that supported spontaneous turn-taking, interactional nuance, and co-constructed meaning (Yusuf et al., 2025). Recent research has highlighted how such affordances are increasingly viable with the integration of AI in immersive environments. Adarkwah et al. (2024) observed that the implementation of ChatGPT within metaverse platforms enhanced linguistic responsiveness, emotional resonance, and contextual adaptability, which resulted in higher perceived immersion and social presence. These findings converge with our observations, where users often escalated participation when avatars produced phrasal irregularities or layered nuance. While dialogic fluidity introduces epistemic risks such as factual drift or alignment drift, it also aligns with dialogic pedagogies that emphasize uncertainty, narrative emergence, and reflexive meaning-making as critical to learning.

Lesson 3: social realism as a byproduct of generative architecture

Realism in virtual environments is a multidimensional construct encompassing visual fidelity, behavioral congruence, and the perceived authenticity of social interaction. While prior research has predominantly focused on avatar (Kim et al., 2023) and behavioral realism (Garcia, 2025), this study foregrounds social realism as an emergent quality enabled by generative AI. In this context, social realism refers to the perceived relational plausibility that emerges when users experience virtual agents as capable of context-sensitive, affectively modulated, and pragmatically coherent interaction. Unlike agency, social realism refers to the phenomenological experience of plausibility in interaction. It is therefore perceptual rather than attributional, and relational rather than structural. In contrast to pre-scripted agents whose communicative range is limited to finite dialogue paths, the generative AI avatars examined here achieved dialogic elasticity that adapted responsively to user inputs across varied scenarios. Perceptions of social authenticity align with the Computers Are Social Actors (CASA) paradigm (Nass et al., 1994), which proposes that users respond to

media agents using social rules when those agents demonstrate simple cues of interactivity (Xu et al., 2022). While CASA was initially developed to explain user responses to rule-based systems, its explanatory power expands in the context of generative AI, where dialogic fluency and adaptive responses intensify the illusion of social presence and intentionality. Crucially, the findings suggest that social realism is not a static design attribute, but a co-constructed phenomenon shaped by the interplay of system responsiveness and user interpretation (Oh et al., 2023). By shifting the analytical lens from appearance and animation to dialogic enactment, this study extends current frameworks and invites a rethinking of how AI-mediated presence is operationalized in educational metaverses.

Lesson 4: instructional intent collides with algorithmic improvisation

Tensions between pedagogical control and technological agency have long characterized educational technology design, but the introduction of generative AI into metaverse learning environments intensifies this friction (Zawacki-Richter et al., 2019). The deployment revealed a structural disjunction between instructional intent and algorithmic improvisation, wherein preconfigured pedagogical goals were frequently destabilized by the stochastic logic of large language models. The AI avatars in this study likewise exhibited discursive plasticity that often exceeded or bypassed designed learning trajectories. This divergence was particularly evident in interactions where expected prompts elicited responses that were contextually plausible but pedagogically tangential. Such outcomes echo broader findings in recent literature. Almeman et al. (2025) emphasized that while AI techniques offer adaptability in virtual environments, they also introduce unpredictability in how instructional materials are navigated. This challenge is compounded in socially oriented metaverse environments where conversational agents are not merely content dispensers but perceived collaborators in meaning-making (Lv, 2023). The system's generative core privileged fluency and topical relevance over domain fidelity, which resulted in moments of interactional drift that blurred the boundaries of instructional authority. These findings underscore the need to revisit design assumptions around control, guidance, and intentionality when integrating non-deterministic AI into learning systems. Rather than treating generative AI as a direct extension of instructional design, future models may need to accommodate improvisation as a structural feature of AI-mediated learning ecologies.

Lesson 5: voice mediation as a threshold for presence and participation

Among the sensory modalities engaged in virtual environments, voice remains uniquely capable of collapsing distance and animating presence through rhythm, tone, and immediacy (Garcia, 2026b; Kojic et al., 2025). The introduction of voice-mediated interaction via integrated text-to-speech and speech-to-text modalities reconfigured the phenomenology of user engagement by establishing a perceptual threshold at which presence was no longer contingent solely on visual fidelity but on auditory synchrony and temporal co-responsiveness. This lesson isolates voice as a modality-specific mechanism that amplifies perceived co-presence by operating not at the level of cognitive attribution (agency) or relational plausibility (realism), but at the sensory threshold of immersion. In contrast to earlier work that locates social presence primarily in avatar realism and proxemics (Latoschik et al., 2017), findings from this study indicate that the fluidity and immediacy of vocal exchange were more salient predictors of perceived authenticity and participatory willingness. The auditory channel functioned as a conduit for verbal content and an amplifier through which latency, intonation, and pacing were interpreted as signals of relational attunement and intentionality. These results align with literature on paralinguistic signaling and cognitive entrainment in human-agent interaction (Schroeder et al., 2006), and are further supported by Kao et al. (2021), who found that self-similar avatar voices significantly increased user immersion and perceived identification in educational games. In this study, users reported heightened co-presence and relational resonance during voice-based interactions even when linguistic content remained repetitive or neutral. Consequently, voice mediation emerged as a threshold mechanism that differentiated passive observation from dialogic co-presence, extending the operational boundaries of participation in AI-augmented immersive environments.

Lesson 6: affective frictions reveal the emotional cost of AI mediation

While generative AI avatars succeeded in approximating linguistic coherence and contextual responsiveness, they revealed pronounced limitations in affective attunement, particularly during moments of emotional ambiguity or interpersonal tension. These limitations gave rise to affective frictions where the avatar's tone, timing, or sentiment failed to resonate with user affect. In contrast to earlier expectations that emotional AI would enhance learner motivation and relational presence, observations from this deployment indicate that affective misalignment can erode trust, stifle engagement, and trigger interpretive dissonance. Similar concerns have been raised in the literature on affective computing in education, where emotional simplification is increasingly recognized as an epistemic and ethical risk (e.g. Ho et al., 2024). While prior studies have documented the utility of AI tools in reducing anxiety and improving relational warmth through empathic signaling and multimodal cues (Hsu et al., 2024; Sargazi Moghadam et al., 2024), the generative AI avatars in this context often failed to deliver affective precision at scale. Users reported that voice tone, pacing, and sentiment occasionally contradicted the intended social register, resulting in interactions that were cognitively valid but emotionally dissonant. Liu et al. (2024) describe this as an emotional support gap, where avatars can deliver content but not the affective congruence needed for learner trust and socio-emotional safety. Within socio-educational metaverses, such misalignments function as diagnostic signals. Designing for affective realism in virtual environments thus demands a recalibration of how emotion, empathy, and engagement are operationalized within the pedagogical interface.

Lesson 7: informal learning thrives in playful and unstructured social contexts

Contrary to assumptions that pedagogical impact resides primarily within structured curricular moments, this deployment revealed that informal and playful activities often became the most generative contexts for learning (Sargazi Moghadam et al., 2024). Within the off-campus environment, AI avatars reached peak social and pedagogical relevance during spontaneous play. These observations resonate with playful constructivism (Marone, 2016), which positions learning as emerging from participatory, situated, and socially immersive activity. In these open-ended interactions, AI avatars served less as tutors and more as social catalysts, redirecting attention, posing incidental questions, or anchoring brief learning exchanges. Lee and Ahn (2025) and Yang et al. (2025) similarly demonstrated how learner agency and active participation were amplified in game-like metaverse settings that allowed for non-linear exploration and collaboration. Their findings further revealed that the presence of narrative-rich activities increased learner persistence and fostered socio-emotional bonding among peers. These results underscore the value of designing educational AI for facilitating relational continuity and improvisational learning (Labadze et al., 2023). Recognizing unstructured and affective contexts as pedagogically meaningful reframes the role of AI avatars as co-participants in socially driven curiosity-centered learning environments. This reframing calls for educational designs that prioritize adaptability, emotional resonance, and playful co-presence over rigid instructional sequencing.

Lesson 8: misalignment between student expectations and AI capabilities

Far from approaching the avatars as experimental tools, many users attributed to them the kind of omnipotent intelligence typically reserved for human instructors or algorithmic oracles. This projection of hyper-competence led to recurring breakdowns in pedagogical alignment, which manifested in user frustration, strategic query reformulation, and moments of disengagement when the avatars failed to meet perceived instructional thresholds. Although several avatars in this deployment were designed primarily for social facilitation rather than instructional delivery, users nonetheless engaged them with expectations of tutoring accuracy or emotional attunement. By routinely overestimating the avatars' capacity for deep content knowledge and emotional resonance, user experienced epistemic dissonance when system responses failed to align with those expectations. These findings echo concerns raised by Kinney et al. (2024), who emphasized the importance of expectation management in sustaining trust and promoting adoption in AI-facilitated environments. Similarly, Castillo-Martínez et al. (2024) observed that inflated perceptions of AI capability can distort learning behaviors and undermine critical reflection. For AI avatars to function

effectively within socio-educational ecosystems, designers must address not only the technical limits of generative architectures but also the cognitive schemas learners bring into the interaction. Anticipating and actively shaping these mental models is essential to closing the gap between system design and learner experience.

Lesson 9: the ethics of social simulation in AI-mediated education

Ethical questions surrounding AI in education are intensifying, particularly as generative systems begin to emulate the relational dimensions of teaching and learning. The deployment of AI avatars designed to simulate empathy, attentiveness, and curiosity invites critical scrutiny when such interactions are perceived as socially authentic but are in fact algorithmically constructed. While these simulations may enhance engagement and perceived presence, they also raise concerns about artificial intimacy and the ethical risks of emotional misrepresentation. Educators in this study expressed discomfort with the implicit deception involved in presenting programmatic responsiveness as genuine social interaction. This issue is echoed in critiques of synthetic relationships in AI design (Park et al., 2024; Williamson & Prybutok, 2024). Nguyen et al. (2023) emphasize that AI-mediated social systems must be held accountable to principles of transparency, human dignity, and relational authenticity, especially when deployed in emotionally charged or developmentally formative contexts. Applying a human-centered AI framework (Calvo et al., 2020), this study underscores that the ethical viability of AI avatars lies not in their technical sophistication but in the intentionality behind their deployment. As avatars increasingly occupy roles associated with care, mentorship, and emotional labor, designers and educators must interrogate what is being simulated, why, and to what pedagogical ends.

Lesson 10: scaffolding AI literacy as a condition for effective interaction

AI literacy emerged as a decisive mediator of engagement quality in this deployment, where meaningful interaction with generative avatars was contingent upon users' conceptual grasp of how such systems operate. Learners who lacked familiarity with the generative mechanisms of AI were more prone to anthropomorphize avatars, misread system outputs, or disengage altogether. Unfortunately, these patterns signaled both epistemic vulnerability and missed pedagogical opportunities (Pinski & Benlian, 2024). These findings align with the assertion that AI literacy is a sociotechnical fluency that includes understanding limitations, ethical implications, and interactional strategies (Walter, 2024). Reinforcing this interpretation is the foundational framework of Long and Magerko (2020), which emphasizes that competencies such as recognizing AI behavior, interrogating its decision-making processes, and discerning the boundary between human and machine agency are essential for learners operating in hybrid learning ecologies. The findings of this study also resonate with the six-construct synthesis articulated by Almatrafi et al. (2024), which foreground the cognitive and ethical sophistication required to critique and co-construct meaning alongside AI systems. Collectively, these insights underscore that without deliberate scaffolding of AI literacy, immersive educational technologies risk amplifying confusion rather than fostering critical engagement.

Synthesizing lessons and implications for human–AI interaction in the metaverse

The lessons synthesized from this study collectively reveal that human–AI interaction in a socio-educational metaverse environment is fundamentally co-constructed, contingent, and shaped as much by learner interpretation as by system architecture (Chiu & Rospigliosi, 2025). AI avatars in the metaverse did not simply function as information conduits, but as relational agents whose generative behaviors provoked shifts in how users perceived agency, presence, and pedagogical authority. This reframing aligns with emerging work that characterizes AI as a distributed cognitive artifact entangled in meaning-making processes (Córdova-Esparza, 2025; Zhang & Siau, 2024). The generative nature of the avatars supported open-ended dialogue and discursive improvisation (Al-Emran, 2024; Yusuf et al., 2025), yet also introduced frictions (e.g. emotional misalignment, interpretive dissonance, and expectation failure) that challenged learners' trust and engagement (Castillo-Martínez et al., 2024; Liu et al., 2024). These tensions were most acute when users lacked AI literacy, which reinforces prior findings that conceptual fluency with AI systems is

foundational to effective participation (Almatrafi et al., 2024; Long & Magerko, 2020; Walter, 2024). At the same time, informal and affectively rich interactions often proved more generative than scripted pedagogical moments, echoing studies on play-based and curiosity-driven learning in immersive learning settings (Sargazi Moghadam et al., 2024; Yang et al., 2025). These findings carry distinct implications for the future of immersive education, with relevance across design, pedagogy, institutional strategy, and research.

Building on these synthesized insights, the following implications outline how AI avatars reshape the expectations, practices, and evaluative frameworks that define educational innovation in the metaverse. For designers, the findings underscore the need to move beyond static behavioral templates toward architectures that accommodate emergent dialogue and user-driven complexity (Adarkwah et al., 2024; Córdova-Esparza, 2025). For educators, the study illuminates both the affordances and tensions introduced by generative AI in learning-adjacent contexts, particularly the necessity of balancing pedagogical intent with algorithmic spontaneity (Lv, 2023; Yusuf et al., 2025). For institutional leaders, the results highlight the strategic importance of adopting evaluative frameworks that support iterative development in high-uncertainty settings where innovation outpaces conventional assessment models (Park et al., 2024; Sargazi Moghadam et al., 2024). Institutions that embed complexity-aware evaluation practices are more likely to sustain innovation cycles and support adaptive system intelligence in real-world deployments (Walter, 2024). For researchers, the study contributes to a growing body of literature that frames AI as an epistemic actor whose integration reshapes design cognition, user experience, and sociotechnical assumptions (Garcia et al., 2023b; Zhang & Siau, 2024). This notion aligns with emerging perspectives that situate AI agents as distributed cognitive artifacts co-constituting the interactional field alongside human users (Liu et al., 2024; Suchman, 2007). Collectively, these insights point to a future in which AI avatars function less as replacements for human agents and more as catalysts for rethinking interaction itself.

Toward a stratified sociotechnical model of human–AI interaction

While the preceding lessons are presented thematically for analytic clarity, they collectively point toward a stratified sociotechnical model of human–AI interaction in the socio-educational metaverse (Figure 4). Rather than viewing AI avatars as isolated technological tools or features, the findings suggest that interaction unfolds across interdependent layers that condition how agency, dialogue, authority, and trust are constructed in situ.

At the foundational level, *technical architecture* structures the possibility space of human–AI interaction. Through large language model (LLM) stochasticity, prompt conditioning, and voice integration systems, AI avatars generate probabilistic responses that shape how users encounter and interpret machine behavior.

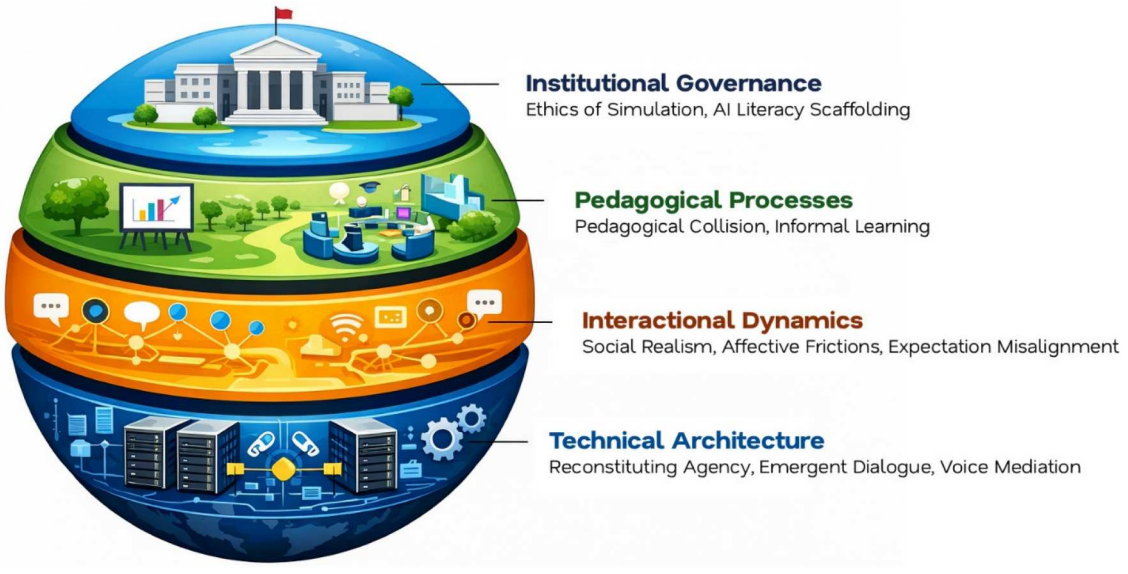

Figure 4. Sociotechnical structure of human–AI interaction in a socio-educational metaverse.

Reconstituted agency and emergent dialogue originate at this foundational level, not as inherent qualities of the avatar but as computational affordances that enable relational interpretation to emerge through human engagement. In this sense, human–AI interaction is technically mediated before it is socially experienced.

Above this substrate, *interactional dynamics* capture how users interpret and respond to avatar behaviors in real time. It is at this level that learners attribute intentionality, perceive social realism, and experience affective frictions or expectation misalignment. Human–AI interaction becomes relational rather than procedural, as users negotiate presence, credibility, and authority through ongoing dialogic exchange. The avatar's generative outputs are transformed into socially meaningful actions through user interpretation, rendering interaction a co-constructed phenomenon rather than a one-sided delivery of information.

The third layer, *pedagogical processes*, reflects how these interactional patterns reshape formal or informal learning structures. When AI avatars engage in unscripted dialogue or discursive improvisation, instructional authority becomes partially distributed across human and artificial agents. Pedagogical collision emerges when algorithmic spontaneity intersects with curricular intent, while informal learning surfaces through affectively rich exchanges that exceed designed lesson boundaries. Here, human–AI interaction reconfigures not only communication patterns but also epistemic roles within the learning environment.

Finally, *institutional governance* situates human–AI interaction within ethical, strategic, and evaluative frameworks. The ethics of simulation and AI literacy scaffolding operate as regulatory mechanisms that shape how avatars are designed, deployed, and interpreted. Institutions mediate the risks and affordances of generative interaction by establishing norms for transparency, accountability, and learner preparedness. Governance therefore conditions how human–AI interaction is legitimized and sustained within educational systems.

Collectively, these layers demonstrate that human–AI interaction in a socio-educational metaverse is not reducible to interface design or conversational quality alone. It is an emergent phenomenon distributed across computational infrastructures, relational enactments, pedagogical reconfigurations, and institutional frameworks. Within this stratified ecology, AI avatars operate as integrative mediators that both embody technical design decisions and activate broader social, educational, and organizational shifts. They translate probabilistic computation into socially legible interaction, while simultaneously reshaping expectations of authority, presence, and participation. As such, AI avatars are not peripheral tools embedded within a digital platform but central actors in the reorganization of how learning, agency, and governance are negotiated in immersive environments. Their role illustrates how technical architectures become consequential forces in the transformation of educational practice and institutional strategy.

Limitations and future research

This study contributes implementation-level insights into the integration of generative AI avatars within a socially situated metaverse environment. However, several conceptual and architectural limitations warrant careful consideration in interpreting these findings.

First, while the study documents dialogic plasticity and emergent interaction patterns, it does not formally model the cognitive mechanisms underlying user interpretation of AI-generated responses. The analysis foregrounds interactional outcomes rather than cognitive load, attribution processes, or mental model formation. As generative AI-driven avatars increasingly simulate naturalistic human interaction (Kohistani et al., 2025), future work must examine how users construct epistemic trust, agency attribution, and intentionality in AI-mediated exchanges.

Second, although structured prompt engineering and moderation constraints were implemented, the study does not systematically evaluate model drift, hallucination persistence, or long-horizon narrative coherence. LLMs are known to exhibit degradation in consistency across extended conversational arcs (Kwan et al., 2024). While field memos documented instances of incoherence, no formal discourse-level analysis was conducted. This limits the ability to quantify narrative stability across multi-session engagements.

Third, the architectural configuration relied on API-based inference without hybrid symbolic layers. Recent literature suggests that VR systems may require modular architectures combining language models with rule-based logic, memory compression, and reinforcement learning to maintain performance in real-time immersive contexts (Özkaya et al., 2025). The present implementation focused on ecological deployment rather than architectural experimentation and thus leaves unresolved questions about optimal hybridization strategies.

Fourth, while voice mediation enhanced perceived realism, the study did not isolate the contribution of multimodal alignment (e.g. synchrony between speech, gesture, and gaze) to immersion outcomes. Prior research on multimodal metaverse highlights the pedagogical and communicative role of multiple interactive cues in shaping user experience within virtual spaces (Garcia, 2026a). The absence of fine-grained multimodal analytics limits interpretability regarding which dimensions of realism most influenced user engagement.

Finally, the DE approach privileges adaptive insight over controlled causal inference. While this strategy aligns with the innovation-stage maturity of AI in metaverse contexts, it constrains the ability to generalize findings across alternative engine architectures, deployment pipelines, or model families. This trade-off reflects underscores the need for subsequent experimental and comparative investigations to strengthen causal claims.

Conclusion

As generative AI becomes increasingly embedded in metaverse ecosystems, longstanding questions around agency, authenticity, and social interaction acquire renewed urgency in both design and pedagogical discourse. While previous studies have examined the instructional affordances of intelligent agents and the immersive qualities of virtual environments, few have investigated the deployment of AI avatars as relational actors within socially situated learning spaces. This study addresses that gap by examining how human–AI interaction unfolds in the metaverse when deterministic, pre-scripted NPCs are replaced by probabilistic, adaptive avatars capable of dialogic improvisation. The results revealed a series of interrelated lessons that collectively illuminate the transformative and disruptive impact of generative AI avatars on educational interaction. These lessons highlight critical themes such as shifting perceptions of agency, the importance of affective realism, the tension between pedagogical structure and algorithmic spontaneity, and the central role of AI literacy in shaping user engagement.

In addition to thematic synthesis, this study contributes in three distinct ways. First, it advances a sociomaterial reframing of AI avatars as attributional and relational phenomena rather than as static instructional tools. Second, it demonstrates the methodological utility of Developmental Evaluation for studying emergent AI-integrated systems in real-world educational deployments, where interpretation and implementation co-evolve. Third, it offers implementation-level design insights that differentiate agency, social realism, voice-mediated presence, and affective friction as analytically distinct yet structurally interdependent dimensions of human–AI interaction. With theoretical and practical implication, the study advances a clearer understanding of generative AI avatars as co-constructors of interactional meaning rather than as passive instructional agents. By surfacing the sociotechnical dynamics that emerge in metaverse-based learning environments, the research contributes to ongoing discourse on how AI systems mediate agency, relationality, and pedagogical intent. Moving forward, critically engaging with the intertwined evolution of generative AI and the metaverse will be essential to shaping immersive learning environments that are adaptive, equitable, and pedagogically meaningful.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Manuel B. Garcia http://orcid.org/0000-0003-2615-422X

References

Acut, D. P., Malabago, N. K., Malicoban, E. V., Galamiton, N. S., & Garcia, M. B. (2025). "ChatGPT 4.0 ghosted us while conducting literature search": Modeling the chatbot's generated non-existent references using regression analysis. *Internet Reference Services Quarterly*, *29*(1), 27–54. https://doi.org/10.1080/10875301.2024.2426793

Adarkwah, M. A., Tlili, A., Shehata, B., Huang, R., Amoako, P. Y. O., & Wang, H. (2024). ChatGPT implementation in the metaverse: Towards another level of immersiveness in education. In Z. Lyu (Ed.), *Applications of generative AI* (pp. 421–436). Springer International Publishing.

Al-Emran, M. (2024). Unleashing the role of ChatGPT in metaverse learning environments: Opportunities, challenges, and future research agendas. *Interactive Learning Environments*, *32*(10), 7497–7506. https://doi.org/10.1080/10494820.2024.2324326

Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open*, *6*, 1–20. https://doi.org/10.1016/j.caeo.2024.100173

Almeman, K., El Ayeb, F., Berrima, M., Issaoui, B., & Morsy, H. (2025). The integration of AI and metaverse in education: A systematic literature review. *Applied Sciences*, *15*(2), 1–35. https://doi.org/10.3390/app15020863

Ansari, A. N., Ahmad, S., & Bhutta, S. M. (2024). Mapping the global evidence around the use of ChatGPT in higher education: A systematic scoping review. *Education and Information Technologies*, *29*(9), 11281–11321. https://doi.org/10.1007/s10639-023-12223-4

Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., … Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. *Computers and Education: Artificial Intelligence*, *8*, 1–15. https://doi.org/10.1016/j.caeai.2025.100407

Bozkurt, A., Xiao, J., Farrow, R., Bai, J. Y. H., Nerantzi, C., Moore, S., … Asino, T. I. (2024). The manifesto for teaching and learning in a time of generative AI: A critical collective stance to better navigate the future. *Open Praxis*, *16*(4), 487–513. https://doi.org/10.55982/openpraxis.16.4.777

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Calvo, R. A., Peters, D., Vold, K., & Ryan, R. M. (2020). Supporting human autonomy in AI systems: A framework for ethical enquiry. In C. Burr, & L. Floridi (Eds.), *Ethics of digital well-being: A multidisciplinary approach* (pp. 31–54). Springer International Publishing.

Castillo-Martínez, I. M., Flores-Bueno, D., Gómez-Puente, S. M., & Vite-León, V. O. (2024). AI in higher education: A systematic literature review. *Frontiers in Education*, *9*, 1–7. https://doi.org/10.3389/feduc.2024.1391485

Chiu, T. K. F., & Rospigliosi, P. (2025). Encouraging human-AI collaboration in interactive learning environments. *Interactive Learning Environments*, *33*(2), 921–924. https://doi.org/10.1080/10494820.2025.2471199

Córdova-Esparza, D.-M. (2025). AI-powered educational agents: Opportunities, innovations, and ethical challenges. *Information*, *16*(6), 1–30. https://doi.org/10.3390/info16060469

Garcia, M. (2026a). Multilingual language learning in a multimodal metaverse: A multidimensional study of communicative, affective, and cognitive development. *Innovation In Language Learning and Teaching*, 1–27. https://doi.org/10.1080/17501229.2026.2621262

Garcia, M. B. (2025). Teachers in the metaverse: The influence of avatar appearance and behavioral realism on perceptions of instructor credibility and teaching effectiveness. *Interactive Learning Environments*, 4334–4350. https://doi.org/10.1080/10494820.2025.2462144

Garcia, M. B. (2026b). The illusion of presence and the reality of engagement: How avatar dynamics define social interaction in an educational metaverse? *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2025.2611127

Garcia, M. B., Adao, R. T., Pempiña, E. B., Quejado, C. K., & Maranan, C. R. B. (2023a). MILES virtual world: A three-dimensional avatar-driven metaverse-inspired digital school environment for FEU group of schools. In *Proceedings of the 7th International Conference on Education and Multimedia Technology*. https://doi.org/10.1145/3625704.3625729

Garcia, M. B., Adao, R. T., Ualat, O. N., & Cunanan-Yabut, A. (2023b). Remodeling a mobile educational metaverse using a co-design approach: Challenges, issues, and expected features. In *Proceedings of the 7th International Conference on Education and Multimedia Technology*. https://doi.org/10.1145/3625704.3625730

Garcia, M. B., Quejado, C. K., Maranan, C. R. B., Ualat, O. N., Adao, R. T., Happonen, A., … Bozkurt, A. (2024). Social relationship development in the metaverse: The roles of embodiment, immersion, and the moderating effect of copresence. In *TENCON 2024 - 2024 IEEE Region 10 Conference (TENCON)*. https://doi.org/10.1109/TENCON61640.2024.10902707

Ho, M.-T., Mantello, P., & Vuong, Q.-H. (2024). Emotional AI in education and toys: Investigating moral risk awareness in the acceptance of AI technologies from a cross-sectional survey of the Japanese population. *Heliyon*, *10*(16), 1–16. https://doi.org/10.1016/j.heliyon.2024.e36251

Hsu, T.-C., Ching, C., & Jen, T.-H. (2024). Artificial intelligence image recognition using self-regulation learning strategies: Effects on vocabulary acquisition, learning anxiety, and learning behaviours of English language learners. *Interactive Learning Environments*, *32*(6), 3060–3078. https://doi.org/10.1080/10494820.2023.2165508

Hwang, A. H.-C., & Won, A. S. (2022). AI in your mind: Counterbalancing perceived agency and experience in human-AI interaction. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491101.3519833

Izquierdo-Álvarez, V., & Jimeno-Postigo, C. (2025). Challenges and opportunities of integrating generative artificial intelligence in higher education: A systematic review. In M. B. Garcia (Ed.), *Pitfalls of AI integration in education: Skill obsolescence, misuse, and bias* (pp. 391–414). IGI Global.

Kao, D., Ratan, R., Mousas, C., & Magana, A. J. (2021). The effects of a self-similar avatar voice in educational games. In *Proceedings of the ACM on Human-Computer Interaction*. https://doi.org/10.1145/3474665

Kim, D. Y., Lee, H. K., & Chung, K. (2023). Avatar-mediated experience in the metaverse: The impact of avatar realism on user-avatar relationship. *Journal of Retailing and Consumer Services*, *73*, 1–11. https://doi.org/10.1016/j.jretconser.2023.103382

Kinney, M., Anastasiadou, M., Naranjo-Zolotov, M., & Santos, V. (2024). Expectation management in AI: A framework for understanding stakeholder trust and acceptance of artificial intelligence systems. *Heliyon*, *10*(7), 1–22. https://doi.org/10.1016/j.heliyon.2024.e28562

Kohistani, A. J., Momand, S., & Zhwak, A. F. (2025). AI driven avatars in virtual reality: A systematic literature review on intelligent agents for enhancing human-computer interaction. *Gameology and Multimedia Expert*, *2*(4), 138–145. https://doi.org/10.29103/game.v2i4.24167

Kojic, T., Vergari, M., Warsinke, M., Ali, D., Möller, S., & Voigt-Antons, J.-N. (2025). Multimodal user experience in extended reality: Exploring hand tracking, voice, and passthrough interactions. In *Proceedings of the 17th International Workshop on IMmersive Mixed and Virtual Environment Systems*. https://doi.org/10.1145/3712677.3720459

Koohang, A., Horn, N. J., Keng-Boon, O., Wei-Han, T. G., Mostafa, A.-E., Cheng-Xi, A. E., … Wong, L.-W. (2023). Shaping the metaverse into reality: A holistic multidisciplinary understanding of opportunities, challenges, and avenues for future investigation. *Journal of Computer Information Systems*, *63*(3), 735–765. https://doi.org/10.1080/08874417.2023.2165197

Kwan, W.-C., Zeng, X., Jiang, Y., Wang, Y., Li, L., Shang, L., … Wong, K.-F. (2024). MT-Eval: A multi-turn capabilities evaluation benchmark for large language models. https://doi.org/10.18653/v1/2024.emnlp-main.1124

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, *20*(1), 1–17. https://doi.org/10.1186/s41239-023-00426-1

Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., & Botsch, M. (2017). The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. https://doi.org/10.1145/3139131.3139156

Lee, S.-M., & Ahn, T. Y. (2025). L2 learner experiences in a playful constructivist metaverse space. *ReCALL*, *37*(1), 129–145. https://doi.org/10.1017/S0958344024000235

Liu, Y., Zhang, H., Jiang, M., Chen, J., & Wang, M. (2024). A systematic review of research on emotional artificial intelligence in English language education. *System*, *126*, 1–13. https://doi.org/10.1016/j.system.2024.103478

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376727

Lv, Z. (2023). Generative artificial intelligence in the metaverse era. *Cognitive Robotics*, *3*, 208–217. https://doi.org/10.1016/j.cogr.2023.06.001

Marone, V. (2016). Playful constructivism: Making sense of digital games for learning and creativity through play, design, and participation. *Journal of Virtual Worlds Research*, *9*(3), 1–18. https://doi.org/10.4101/jvwr.v9i3.7244

Miles, M. B., Huberman, A. M., & Saldana, J. (2013). *Qualitative data analysis: A methods sourcebook*. SAGE Publications. https://books.google.com.ph/books?id=p0wXBAAAQBAJ

Miranda, J. P. P., Cruz, M. A. D., Fernandez, A. B., Balahadia, F. F., Aviles, J. S., Caro, C. A., … Gaña, E. P. (2025). Erosion of critical academic skills due to AI dependency among tertiary students: A path analysis. In M. B. Garcia (Ed.), *Pitfalls of AI integration in education: Skill obsolescence, misuse, and bias* (pp. 25–48). IGI Global.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/191666.191703

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, *28*(4), 4221–4241. https://doi.org/10.1007/s10639-022-11316-w

Oh, H. J., Kim, J., Chang, J. J. C., Park, N., & Lee, S. (2023). Social benefits of living in the metaverse: The relationships among social presence, supportive interaction, social self-efficacy, and feelings of loneliness. *Computers in Human Behavior*, *139*, 1–11. https://doi.org/10.1016/j.chb.2022.107498

Onu, P., Pradhan, A., & Mbohwa, C. (2024). Potential to use metaverse for future teaching and learning. *Education and Information Technologies*, *29*(7), 8893–8924. https://doi.org/10.1007/s10639-023-12167-9

Ortlipp, M. (2008). Keeping and using reflective journals in the qualitative research process. *The Qualitative Report*, *13*(4), 695–705. https://doi.org/10.46743/2160-3715/2008.1579

Özkaya, S., Berrezueta-Guzman, S., & Wagner, S. (2025). How LLMs are shaping the future of virtual reality. *IEEE Access*, *13*, 193335–193355. https://doi.org/10.1109/ACCESS.2025.3631594

Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. *Patterns*, *5*(5), 1–16. https://doi.org/10.1016/j.patter.2024.100988

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. The Guilford Press. https://psycnet.apa.org/record/2010-19231-000

Pinski, M., & Benlian, A. (2024). AI literacy for users – A comprehensive review and future research directions of learning methods, components, and effects. *Computers in Human Behavior: Artificial Humans*, *2*(1), 1–22. https://doi.org/10.1016/j.chbah.2024.100062

Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, *14*(1), 29–48. https://doi.org/10.1177/1356389007084674

Sargazi Moghadam, T., Ali, D., Mansoureh, D., & Mashayekh, S. (2024). Effective principles of informal online learning design: A theory-building metasynthesis of qualitative research. *Interactive Learning Environments*, *32*(7), 3665–3685. https://doi.org/10.1080/10494820.2023.2188398

Schroeder, R., Heldal, I., & Tromp, J. (2006). The usability of collaborative virtual environments and methods for the analysis of interaction. *Presence: Teleoperators and Virtual Environments*, *15*(6), 655–667. https://doi.org/10.1162/pres.15.6.655

Suchman, L. A. (2007). *Human-Machine reconfigurations: Plans and situated actions*. Cambridge University Press. https://psycnet.apa.org/record/2007-00090-000

Tlili, A., Huang, R., Shehata, B., Liu, D., Zhao, J., Metwally, A. H. S., … Burgos, D. (2022). Is metaverse in education a blessing or a curse: A combined content and bibliometric analysis. *Smart Learning Environments*, *9*(1), 1–31. https://doi.org/10.1186/s40561-022-00205-x

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, *21*(1), 1–29. https://doi.org/10.1186/s41239-024-00448-3

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, *252*, 124167. https://doi.org/10.1016/j.eswa.2024.124167

Williamson, S. M., & Prybutok, V. (2024). The era of artificial intelligence deception: Unraveling the complexities of false realities and emerging threats of misinformation. *Information*, *15*(6), 1–43. https://doi.org/10.3390/info15060299

Xiao, J., Bozkurt, A., Nichols, M., Pazurek, A., Stracke, C. M., Bai, J. Y. H., … Themeli, C. (2025). Venturing into the unknown: Critical insights into grey areas and pioneering future directions in educational generative AI research. *TechTrends*, *69*(3), 582–597. https://doi.org/10.1007/s11528-025-01060-6

Xiaoyu, W., Zamzami, Z., & Hai Leng, C. (2025). Generative artificial intelligence in pedagogical practices: A systematic review of empirical studies (2022–2024). *Cogent Education*, *12*(1), 1–21. https://doi.org/10.1080/2331186X.2025.2485499

Xu, K., Chen, X., & Huang, L. (2022). Deep mind in social responses to technologies: A new approach to explaining the computers are social actors phenomena. *Computers in Human Behavior*, *134*, 1–13. https://doi.org/10.1016/j.chb.2022.107321

Yang, E., Renard, D., & Chollet, A. (2025). Onboarding for a new playful narrative adventure in game metaverses. *Technological Forecasting and Social Change*, *213*, 1–12. https://doi.org/10.1016/j.techfore.2025.123999

Yusuf, H., Money, A., & Daylamani-Zad, D. (2025). Pedagogical AI conversational agents in higher education: A conceptual framework and survey of the state of the. *Educational Technology Research and Development*, *73*(2), 815–874. https://doi.org/10.1007/s11423-025-10447-4

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0

Zhang, Y., & Siau, K. (2024). Meta-Entrepreneurship: An analysis theory on integrating generative AI, agentic AI, and metaverse for entrepreneurship. *Journal of Global Information Management*, *32*(1), 1–21. https://doi.org/10.4018/JGIM.364094

Appendix A. Triangulated data contribution matrix

Lesson	Interviews	Teacher Reflections	User Feedback Tools	Field Memos	Documentation
Reconstituting Agency	S – Reframed avatar perception	M – Surprised by AI responsiveness	M – AI felt real during chats	S – Memo on user shift in language	S – Design goals re: agency
Emergent Dialogue	S – Described natural dialogues	S – Noted meaningful back-and-forth	S – Open-ended talk appreciated	S – Described fluid discourse	-
Social Realism	S – Attributed human traits to AI	S – Felt AI had "personality"	S – Unexpected realism noted	M – "Spoke like a peer" noted	-
Pedagogical Collision	S – Raised instructional mismatch	M – Frustration with AI misfires	M – Wanted clearer guidance	S – Documented design-intent clash	S – Logs showed divergence patterns
Voice Mediation	M – Noted impact of voice	M – Noted voice boosts attention	S – Voice = presence	-	-
Affective Frictions	M – Identified emotional mismatch	S – Called tone "off" or "weird"	S – Lacked emotional nuance	S – Log of awkward sentiment exchange	-
Informal Learning	S – Praised playful learning	S – Spontaneous learning preferred	S – Enjoyed camp more than lessons	M – Camp play led to insight	-
Expectation Misalignment	S – Misaligned expectations	M – Overestimated AI capacity	S – Believed AI was "like Google"	-	-
Ethics of Simulation	S – Ethical discomfort discussed	M – Questioned emotional authenticity	-	S – Flagged discomfort in logs	-
AI Literacy Scaffolding	S – Linked understanding to outcomes	S – Called for AI explanation	S – Confused about limits	S – Misconceptions observed	S – AI literacy as dev goal

Note: S = Strong contribution: The theme emerged directly or was significantly discussed in this data source. M = Moderate contribution: The theme was mentioned or supported, but not central to that data stream. A blank cell means no substantial reference to the theme was found in that source.

Appendix B. Sample feedback and observed impact

Lesson	Sample Feedback	Observed Impact
Reconstituting Agency	- "It felt like the avatar actually had intentions. It didn't feel like just a help bot." - "After a few minutes I forgot it was AI. It responded like a real person." - "The responses were so natural it made me wonder if someone was watching or guiding it."	Increased trust and willingness to engage in longer conversations; improved relational presence; sparked questions about AI authenticity.
Emergent Dialogue	- "The AI asked follow-up questions that clearly weren't scripted. That surprised me in a good way." - "Sometimes the conversation went in a completely unexpected direction." - - "It didn't always give a clear answer, but I liked how it flowed like a real chat."	Supported improvisational learning; encouraged curiosity but sometimes caused confusion among users expecting structure.
Social Realism	- "I laughed when it joked back at me. It actually made me feel like it had a sense of humor." - "The AI mentioned a dog story after I talked about pets. That made the whole interaction feel more real." - "I was convinced it remembered something I said earlier, which was impressive."	Strengthened emotional engagement; made interactions feel authentic and socially rich; created an illusion of memory or continuity.
Pedagogical Collision	- "I asked something from the lesson, and it answered with something unrelated. It wasn't helpful at that time." - "The avatar used a metaphor when I just needed a straight answer." - "I liked how it responded, but I had no idea what the takeaway was supposed to be."	Reduced instructional clarity; distracted students from learning goals; highlighted conflicts between creative dialogue and structured content.

(*Continued*)

Continued.

Lesson	Sample Feedback	Observed Impact
Voice Mediation	- "Hearing it speak was more engaging than reading any text. It made me pay closer attention." - "Even when it repeated things, the voice made it feel more present and real." - "I answered more quickly when it used voice. It helped me stay focused."	Heightened engagement through auditory feedback; increased sense of co-presence and responsiveness; encouraged verbal participation.
Affective Frictions	- "It said "I understand" in a really upbeat tone while I was being serious. That threw me off." - "The avatar tried to comfort me, but the timing didn't feel right." - "After a while, I stopped talking to it. It didn't seem to pick up on my mood."	Caused emotional disconnect; reduced trust in avatar empathy; limited emotional disclosure and interaction depth.
Informal Learning	- "I learned more from the kayaking game than any of the guided sessions." - "We started talking about my favorite movie during a race. That ended up being surprisingly educational." - "The fun activities made it easier to talk. Learning just kind of happened."	Reinforced the value of playful, unstructured settings; encouraged curiosity-led learning; boosted engagement outside formal tasks.
Expectation Misalignment	- "I thought it would act like a tutor. It didn't meet that expectation." - "Some of us were frustrated when it couldn't give a straight answer." - "At first, I assumed it was super smart. Later I realized it's still limited."	Led to user frustration and unrealistic expectations; prompted reflection on AI limitations; demonstrated the need for upfront guidance.
Ethics of Simulation	- "It acted like it cared, but I knew it wasn't real. That made me feel weird." - "The friendliness was nice, but kind of fake once I thought about it." - "It simulated empathy well. But after a while, I didn't trust it."	Raised ethical concerns about simulated emotions; created discomfort around authenticity; highlighted risks of emotional manipulation.
AI Literacy	- "Some students thought the AI was supposed to know everything. That made things confusing when it didn't." - "Once I learned how it worked, I adjusted my questions and got better results." - - "It would've helped to get a quick overview of what the AI can actually do."	Improved interaction quality when AI limitations were understood; reduced frustration; emphasized the importance of AI onboarding.