Received: 27 November 2024 | Accepted: 25 May 2025

DOI: 10.1111/bjet.13607

ORIGINAL ARTICLE

British Journal of Educational Technology BERA

Human–AI collaborative learning in mixed reality: Examining the cognitive and socio-emotional interactions

Belle Dang[1] | Luna Huynh[1] | Faaiz Gul[1] | Carolyn Rosé[2] | Sanna Järvelä[1] | Andy Nguyen[1]

[1]Learning and Educational Technology (LET) Research Lab, University of Oulu, Oulu, Finland

[2]Human–Computer Interaction Institute Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Correspondence
Belle Dang, Learning and Educational Technology (LET) Research Lab, University of Oulu, Oulu, Finland.
Email: belle.dang@oulu.fi

Funding information
Research Council of Finland, Grant/Award Number: 350249; Research Council of Finland - University of Oulu profiling project Profi7 Hybrid Intelligence, Grant/Award Number: 352788

The rise of generative artificial intelligence (GAI), especially with multimodal large language models like GPT-4o, sparked transformative potential and challenges for learning and teaching. With potential as a cognitive offloading tool, GAI can enable learners to focus on higher-order thinking and creativity. Yet, this also raises questions about integration into traditional education due to the limited research on learners' interactions with GAI. Some studies with GAI focus on text-based human–AI interactions, while research on embodied GAI in immersive environments like mixed reality (MR) remains unexplored. To address this, this study investigates interaction dynamics between learners and embodied GAI agents in MR, examining cognitive and socio-emotional interactions during collaborative learning. We investigated the paired interactive patterns between a student and an embodied GAI agent in MR, based on data from 26 higher education students with 1317 recorded activities. Data were analysed using a multi-layered learning analytics approach, including quantitative content analysis, sequence analysis via hierarchical clustering and pattern analysis through ordered network analysis (ONA). Our findings identified two interaction patterns: type (1) AI-led Supported Exploratory Questioning (AISQ) and type (2) Learner-Initiated Inquiry (LII) group. Despite their distinction in characteristic, both types demonstrated comparable levels of socio-emotional engagement and exhibited meaningful cognitive engagement, surpassing the superficial content reproduction that can be observed in interactions with GPT models. This study contributes to the human–AI collaboration and learning studies, extending understanding to learning in MR environments and highlighting implications for designing AI-based educational tools.

KEYWORDS

embodied agent, generative artificial intelligence, human–AI collaboration, mixed reality, ordered network analysis, social interactions

Practitioner notes

What is already known about this topic

- Socio-emotional interactions are fundamental to cognitive processes and play a critical role in collaborative learning.
- Generative artificial intelligence (GAI) holds transformative potential for education but raises questions about how learners interact with such technology.
- Most existing research focuses on text-based interactions with GAI; there is limited empirical evidence on how embodied GAI agents within immersive environments like Mixed Reality (MR) influence the cognitive and socio-emotional interactions for learning and regulation.

What this paper adds

- Provides first empirical insights into cognitive and socio-emotional interaction patterns between learners and embodied GAI agents in MR environments.
- Identifies two distinct interaction patterns: AISQ type (structured, guided, supportive) and LII type (inquiry-driven, exploratory, engaging), demonstrating how these patterns influence collaborative learning dynamics.
- Shows that both interaction types facilitate meaningful cognitive engagement, moving beyond superficial content reproduction commonly associated with GAI interactions.

Implications for practice and/or policy

- Insights from the identified interaction patterns can inform the design of teaching strategies that effectively integrate embodied GAI agents to enhance both cognitive and socio-emotional engagement.
- Findings can guide the development of AI-based educational tools that capitalise on the capabilities of embodied GAI agents, supporting a balance between structured guidance and exploratory learning.
- Highlights the need for ethical considerations in adopting embodied GAI agents, particularly regarding the human-like realism of these agents and potential impacts on learner dependency and interaction norms.

INTRODUCTION

The rapid advancement and widespread adoption of generative artificial intelligence (GAI) have brought transformative changes across various sectors, including education (Cukurova, 2025; Giannakos et al., 2024). This technology demonstrates remarkable capabilities in generating human-like content, such as text, images and videos. However, these advancements have raised immediate concerns regarding academic integrity because distinguishing between students' original work and content generated by GAI has become increasingly challenging (Xia et al., 2024). A deeper and more troubling issue is the potential long-term impact on learners' critical thinking and cognitive processes due to reliance on GAI. For example, Sandhaus et al. (2024) found that students in a human-computer interaction education and technology design course report their tendency to over-rely on using GAI without thinking critically about their approach and choices. Therefore, efforts to address these challenges have proven complex (Giannakos et al., 2024; Hamilton et al., 2023). It is technically impractical to impose outright bans on GAI. Furthermore, the integration of GAI into professional environments underscores the urgent need for educational systems to prepare learners with the skills and agency necessary to navigate and thrive in contexts enhanced by GAI.

While GAI poses challenges in terms of academic integrity and diminishes learners' agency, it also offers significant opportunities to support and enhance human learning (Järvelä et al., 2023; Weber et al., 2023). For instance, GAI can provide immediate feedback in response to learners' needs (Escalante et al., 2023), offer scalable personalised learning options (Naik et al., 2025) and stimulate learning processes through interactive engagement (Owoseni et al., 2024). Recent research highlights the importance of developing evidence-based strategies to integrate GAI effectively and responsibly into educational contexts, leveraging its potential while addressing associated risks (Nguyen, Kremantzis, et al., 2024). Despite the growing interest, much of the current literature on GAI in education remains speculative, with limited empirical studies examining its tangible implications for changes in learner behaviours and the processes involved in learning.

Recent studies have attempted to address this gap by providing preliminary evidence of GAI's applications in different learning contexts, such as academic writing (eg, Nguyen, Hong, et al., 2024; Weber et al., 2023), instructional design (eg, Hodges & Kirschner, 2024) and other areas (eg, Cooper, 2023). For example, Cooper (2023) found that while GAI tools can assist educators with tasks, such as creating lesson plans and rubrics, their outputs often lack critical evidence, emphasising the need for users to engage critically with these tools. Still, most of these works focus on the potential of GAI rather than providing a grounded understanding of its tangible impacts on learning (Hennessy et al., 2024). One of the first few attempts to look in this dimension is Nguyen, Hong, et al. (2024), who found that writing interactions characterised by active engagement and critical use of GAI, rather than passive reliance, led to deeper learning experiences. This highlights the importance of understanding learners' interactions with GAI, as its effect on learning is closely tied to the nature and quality of these interactions. This finding aligns with the previously conceptualised phenomenon of human–AI collaboration in learning (Järvelä et al., 2023; Molenaar, 2022) and highlights the need to investigate how learners can effectively interact with GAI to enhance learning.

What is less clear is the nature of interactions for human–AI collaboration between learners and GAI in the MR environment, where GAI can assume an embodied form capable of immersive, human-like multimodal interactions. According to contemporary learning theories, social interactions or socio-emotional interactions as referred to in this study, are fundamental to cognitive development, serving both as a motivator and as a precursor to learning (Jeong et al., 2022). Socio-emotional interactions are widely regarded as essential to the learning process, particularly in their capacity to promote self-regulation and shared regulation in collaborative learning environments (Isohätälä et al., 2017). They shape the reciprocal interactions tied to emotion, motivation and participation, enabling learners to refine perspectives, regulate learning and deepen cognitive development and engagement. This raises the question of whether GAI can be seen as a collaborative learning partner driving cognitive and importantly, socio-emotional engagement and not just merely a tool for information retrieval. In this study, engagement in these dimensions refers to the frequency and depth of cognitive or socio-emotional interactions, highlighting the extent of learners' active participation in reasoning, problem-solving (cognitive) or relational, empathetic and supportive exchanges (socio-emotional). Embodied GAI, including holographic agents, avatars or social robots (eg, Wolf et al., 2022), has the potential to engage learners through multimodal communication (eg, Hong et al., 2021), incorporating voice, gestures and spatial interactions. This capability allows embodied GAI to create immersive learning experiences that mimic real-life social interactions. Researchers generally suggest that augmenting the human-like attributes of embodied AI in the form of robots can significantly improve learners' emotional engagement, attentiveness and instructional outcomes (Kim et al., 2022; Li et al., 2023), but their adoption is constrained by high costs and limited scalability. By blending virtual and physical elements, holographic embodied GAI agents in MR environments instead offer an adaptive, cost-effective and scalable platform for embedding into educational contexts. Such settings enable personalised learning experiences while fostering the cognitive and socio-emotional interactions that are often missing in other forms (ie, text-based) of human–AI collaboration. Accordingly, this study aims to investigate the cognitive and socio-emotional interactions between human learners and embodied GAI within mixed reality (MR) environments through the following research questions:

RQ1: What are the key patterns of cognitive and socio-emotional interactions in human–AI collaborative learning with an embodied GAI agent in MR?
RQ2: How do the structural and temporal characteristics of these key patterns in human–AI collaborative learning differ with respect to the cognitive and socio-emotional aspects?

THEORETICAL BACKGROUND

Learners' cognitive and socio-emotional interactions in learning

Research has demonstrated that cognitive and socio-emotional interactions significantly influence learning outcomes (Kwon et al., 2014). Cognitive development occurs within the context of social interactions, where learners engage in dialogue and shared understanding, co-constructing knowledge through social and collaborative processes (Vygotskii, 1978). Building on this foundational perspective, contemporary research emphasises the complementary role of emotional regulation, empathy and interpersonal skills in fostering collaboration and deeper cognitive engagement (Bakhtiar et al., 2018; Järvenoja & Järvelä, 2009; Webb & Mastergeorge, 2003). Cognitive interactions, such as reasoning, problem-solving and knowledge construction, are often shaped and enriched by socio-emotional dynamics, including communication, motivation and shared goals (Higgins, 1999; Isohätälä et al., 2020). For instance, effective emotional support and interpersonal understanding can reduce anxiety and enhance cognitive processing by fostering a positive learning environment (Lee & Reeve, 2020). Such relational and emotional exchanges, referred to as socio-emotional interactions, have garnered growing research interest and empirical support for their critical role in fostering effective collaboration (Huang & Lajoie, 2023), knowledge co-constructions and deeper learning. Together, these insights underscore that learning is a multifaceted process, intertwining cognitive, emotional and social dimensions. Understanding this interplay is crucial for designing learning experiences that promote deeper engagement and more meaningful outcomes.

Despite the growing recognition of the importance of cognitive and socio-emotional interactions, challenges remain in promoting these interactions effectively. One major challenge is facilitating these interactions in classrooms, particularly large or culturally diverse ones, where teachers' and individual attention is limited, and differences in communication styles or cohesion perceptions can impede collaboration (Kuo et al., 2015; Zamecnik et al., 2024). To overcome these challenges, research emphasises the role of different forms of learning regulation (self-, co- and socially shared-) in helping learners plan, monitor, evaluate and manage their cognitive activities and socio-emotional interactions effectively (Järvelä & Hadwin, 2024; Winne & Hadwin, 1998). Supporting learners' regulation process is thus critical but often difficult, as without appropriate scaffolding, many learners often lack the necessary skills to do so (Järvelä et al., 2021). In this aspect, interventions focusing on both cognitive scaffolding and socio-emotional training have shown promise (Näykki et al., 2021). Recently, the use of AI-driven collaborative tools, such as conversational agents, offers new ways to support learners in monitoring and regulating their interactions (Edwards et al., 2025), allowing for human–AI collaboration for knowledge co-construction (Lee et al., 2024). In this essence, GAI and MR technologies offer potential by creating interactive environments where AI agents engage learners through human-like multimodal interactions. GAI supports cognitive regulation with immediate feedback and adaptive challenges, while MR enables embodied AI agents to aid socio-emotional regulation by responding to learners' emotional cues empathetically.

Human–AI shared regulation in learning theoretical framework

Building upon the potential of GAI and MR technologies to support cognitive and socio-emotional processes for learning and regulation, it becomes essential to investigate the underlying mechanisms that make this collaboration effective. With the rapid advancement of GAI and its growing influence on learning, recent studies also further emphasise the importance of investigating collaboration patterns between humans and AI to better inform and support learning processes (Lee et al., 2024; Nguyen, Hong, et al., 2024). In particular, this research highlights the need to understand how human–AI interactions can enhance learning, focusing on the dynamics that foster effective collaboration (Nguyen, Hong, et al., 2024). Drawing from well-established self-regulated learning (SRL) theory (Winne & Hadwin, 1998) and the concept of triggers for shared regulation of learning (Järvelä & Hadwin, 2024), the Human–AI Shared Regulation in Learning (HASRL) model introduced by (Järvelä et al., 2023) offers a framework of collaborative learning where both human learners and AI systems actively co-regulate the learning process. This framework integrates learners' self-regulatory skills with AI's adaptive feedback mechanisms, establishing a reciprocal relationship. AI, rather than being a passive tool, functions as an active co-regulator, contributing to shared metacognitive strategies. Learners benefit from real-time interactions and feedback provided by GAI, while their responses and engagement continuously inform the AI's adaptive processes. This dynamic interaction fosters a collaborative learning environment in which both human and AI contributions are critical for improving educational outcomes for increasingly popular AI-integrated environments. This study adopts the HASRL model as the theoretical framework for examining human–AI collaborative learning with embodied GAI in MR environments.

AI agents in mixed reality learning contexts

Extended reality (XR) technologies have gained increasing popularity in education and training, particularly in fields, such as medical training and STEAM education (Bower et al., 2017; Chang et al., 2023; Maas & Hughes, 2020). XR offers immersive, risk-free and repeatable environments that can enhance performance in a wide range of tasks. XR encompasses virtual reality (VR), augmented reality (AR) and MR. While AR overlays digital content onto the real world, enriching the user's perception without replacing their physical surroundings, VR provides a fully immersive digital environment. MR, positioned between AR and VR, integrates physical and digital elements in a way that allows users to interact with both simultaneously. In MR environments, virtual and physical domains are spatially aligned, enabling seamless interaction.

The integration of artificial intelligence (AI) agents into XR environments is emerging as a promising approach to addressing challenges and unlocking opportunities for cross-development between AI and XR. For example, Liaw et al. (2023) proposed an AI-enabled VR simulation (AI-enabled VRS) to assess nursing students' competencies and experiences in communicating with an AI-powered medical doctor. However, the majority of existing research has focused on AI agents within VR-based learning environments. The concept of AI agents in education is not new, stretching back decades from the earliest forays into computer-based teaching (Carbonell, 1970), to broader explorations of MR agents (Holz et al., 2011). The rapid advancement of MR technologies, particularly with the development of MR head-mounted displays (HMDs), has introduced new possibilities for teaching and learning. Despite these advancements, our review of the existing literature indicates a scarcity of studies exploring the use of embodied AI agents in MR HMD environments for educational purposes. Most studies on embodied agents in education broadly defined any virtual body representation as 'embodiment' (Jeon et al., 2023). This leaves most examined embodied agents as chatbots with 2D or 3D human avatars on flat screens (Wang et al., 2024; Yang et al., 2022) rather than the true spatial embodiment that MR representation can afford. Meanwhile, earlier MR agent research featuring stronger virtual embodiments typically employed a weaker agent architecture (Holz et al., 2011), likely constrained by the limitations of the technologies of the time, both in terms of GAI models and MR hardware. As a result, this led to only partial or conceptual explorations of how such agents might enhance learning experiences, rather than providing tangible evidence of their impact on specific learning processes. Our study aimed to address this gap by leveraging recent advances in GAI and MR technologies to investigate how embodied GAI agents in MR HMDs can support learning and teaching processes, contributing to the growing body of knowledge in this field.

METHODS

Participant and procedure

Data collection involved 26 higher education students (8 males and 18 females; 22 masters' and 4 doctoral students) enrolled in English-medium international programmes at a Finnish university. The cohort includes both international and Finnish students, all of whom demonstrated good command of English skills required for academic study. All conversations were conducted in English. These students engaged in paired interactive learning dialogues with KAI, a GAI-based speech interface, to explore AI ethics using a Quest3 headset in a controlled setting. The learning task, designed by learning science researchers for a higher education context, involved open-ended dialogue with KAI and exploration of embedded case studies and materials. KAI was designed to foster bidirectional communication by embedding open-ended questions within its responses. These aimed to support exploration, understanding and viewpoint development, concluding with an informal oral question to express what they had learned and reflected upon. The task was structured to promote and assess ethical reasoning, critical reflection and perspective-taking. The learning environment used the Meta Quest 3's passthrough MR mode to integrate digital content seamlessly into the physical environment, rather than merely overlaying it. Real-time spatial mapping enabled the system to detect the layout and geometry of the physical space and anchor digital elements accordingly, ensuring they occupied realistic, collision-free positions. Participants freely navigated this blended space and used natural hand gestures to grab, reposition or activate spatially anchored digital elements (eg, video panels and instructions). While explicit interaction with real-world objects was not the focus of the current task design, the system tracked learners' physical positioning and head orientation, allowing the KAI agent to continuously adjusted its body posture and gaze direction in response to participants' real-time positioning, sustaining face-to-face alignment. This bidirectional responsiveness to both learner movement and environmental structure reflects core characteristics of MR, in which digital content is expected to detect, respect and respond dynamically to the real-world setting and the learner's embodied presence within it (Flavián et al., 2019; Kara & Çakıcı Alp, 2024). The MR setup enabled free movement and embodied engagement in real space, supporting immersion without the disorientation typical of fully virtual environments. This offered a natural setting well-suited for dialogue-based learning and human-like conversational interaction with KAI. Details on the development of the learning environment and embodied agent can be found in Appendix A. The topic on AI ethics was chosen as it provides a meaningful context for exploring GAI implications, aligning with the study's focus on human–AI collaboration, while also stimulating critical thinking and reflective dialogue on a real-world issue. The study was conducted in LeaF Research Infrastructure (http://www.oulu.fi/leaf-eng). Before the interaction, participants completed a pre-questionnaire to provide background information. To assess their perceptions of KAI as a GAI-based speech interface, the Partner Model Questionnaire (PMQ) (Doyle et al., 2023) was administered both before and after the session. In the pre-survey, learners were asked about their general perceptions of GAI-based speech interfaces, drawing from prior experiences with systems, such as Siri, Alexa, Google Assistant or GPT, or their expectations if they lacked such experience. The post-survey, in contrast, specifically evaluated learners' impressions of KAI after their direct interaction with it during the session. The session, capped at 30 minutes to prevent headset discomfort, was divided into three phases: (1) a 5-minute orientation to the virtual environment and gesture controls, (2) a 20-minute interactive discussion with KAI on ethical AI topics, and concluded with (3) a 5-minute oral assessment.

The experiment followed a structured protocol designed to ensure consistency and high-quality data collection. After consenting, each participant was briefed on the task expectations and session structure before being guided to the designated setup area (see Figure 1). Researchers assisted participants in wearing the Quest3 headset and attaching a wireless microphone to capture audio with minimal interference. The setup included two TP-LINK Tapo C225 video cameras positioned to record third-person views of participants' full-body movements, supplemented by the Quest3's built-in first-person camera for a comprehensive view of interactions.

Data analysis

By examining patterns of cognitive and socio-emotional interactions in learner-GAI collaborative learning, we focused on two main areas: (1) how learners and the GAI agent engaged in cognitive interactions for knowledge construction (eg, questioning and reasoning),

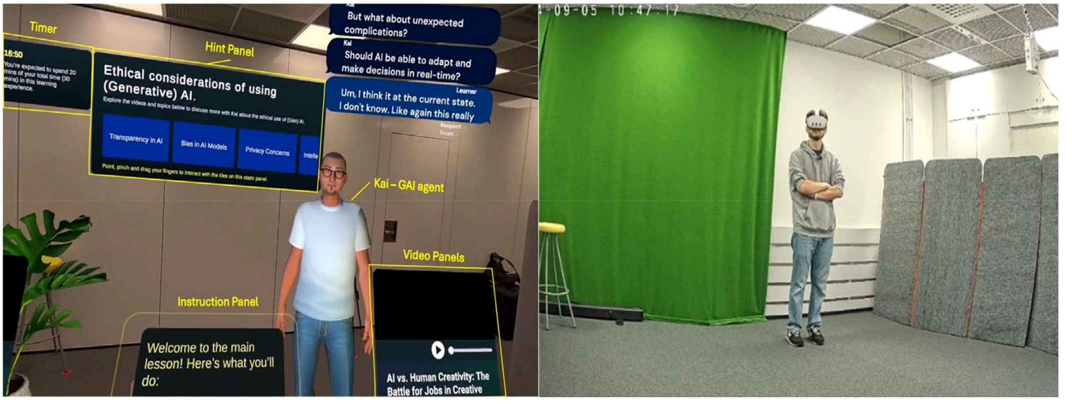


FIGURE 1 Design and context of the experiment.

and (2) the interplays of socio-emotional interactions (eg, expressions of support, social pleasantries or personal emotion) with these cognitive interactions. These patterns are key to understanding learner engagement with GAI in MR environments. We adopted Ouyang et al.'s (2023) AI-driven learning analytics approach, which combines multiple techniques suited for such process-oriented data and has proven effective in similar contexts (Dang et al., 2024; Nguyen, Hong, et al., 2024). First, quantitative content analysis (QCA) is applied to learner-GAI discussion transcripts to code each interaction turn for both learner and GAI for interactions related to knowledge construction strategies and socio-emotional interaction. The second layer examines the similarity of the pair's sequences in order to detect distinct patterns of engagement. Finally ordered network analysis (ONA) is used to explore the regularities and interaction patterns within the identified clusters. Additionally, to examine changes in learners' perceptions of the GAI as a collaborative partner, PMQ scores were analysed using a Paired-sample Wilcoxon Signed Rank Test to assess differences between pre- and post-sessions. Rank-biserial correlations were calculated as effect sizes to evaluate the strength and direction of these changes.

Quantitative content analysis

Quantitative content analysis was conducted on the dialogue transcripts from the main 20-minute session, where learners discussed AI ethics with GAI in the MR environment. Given that each partner's utterance often served multiple discussion purposes, the unit of analysis was defined as the student's or GAI's *unit of meaning* to ensure objectivity. This unit could represent either a complete utterance or a meaningful segment within it. Each meaningful unit was assigned a code corresponding to either cognitive or socio-emotional interactions aspect (see Table 1). The coding scheme drew on Shukor et al.'s (2014) coding scheme on cognitive interactions for knowledge construction strategies, and Kauffeld et al.'s (2018) for socio-emotional interactions. These constructs are critical within the fields of learning sciences, human–AI interaction and shared regulation, providing a robust foundation for analysing the cognitive and socio-emotional interactions within the dialogue.

Two raters conducted reliability checks for the coding procedure. Raters 1 and 2 independently coded an initial 10% of the dataset using the proposed scheme, achieving Cohen's Kappa $\kappa = 0.76$ (Cohen, 1960), with code-specific scores between 0.71 and 0.84, except Verify 0.49. A brief discussion was held to refine interpretation before independently

TABLE 1 Coding scheme for analysing cognitive and socio-emotional interactions.

Categories	Codes	Definitions	Examples
Cognitive interactions			
Asking questions	AskSimple	Asking questions that do not require an explanation (facts or simple questions)	Sorry, can you repeat the proposed solution? Baba? Sorry, could you clarify what you mean by baba?
	AskExplain	Asking questions that require an explanation (comprehension or elaboration)	Do you think an AI system should be able to diagnose illnesses? Why or why not?
	Verify	Verification or asking for agreement	AI: […] let me clarify, are you asking about the role of big corporations in the context of AI and human creativity?
Giving answers	AnsSimple	Answering without explanation	Okay, it means Philip. I have just seen 1 video
	AnsExplain	Answering with explanation (using arguments or by asking a counter-question)	Um, I think it's possible. Yet again, I think like for example, if it's like a predetermined action, if there's no spontaneity required, then, um, I think it's okay. As long as like the plan is also decided by a human doctor
Giving information	InfoSimple	Giving information (an idea or thought) without elaboration	Sure. We could probably discuss this in, uh, in the practice instead of this. I did not watch it but I think that is an interesting topic
	InfoExplain	Giving information (an idea or thought) with elaboration	When it comes to the financial side of things for example. You could consider making it illegal to make money off of any AI-generated content that has been trained on […]
	ReferEarlier	Referring to earlier remark/information	The earlier ideas. As you said about the paintings, you know, artists can explore much more variety of techniques or ideas[…]
	EvaluateSum	Evaluating the content (summarising/concluding)	Interesting idea! A watermark could serve as a form of disclosure, helping to differentiate AI-generated content from original work
	AgreeSimple	Accepting contribution of another participant without elaboration	Hm. Maybe Sounds good. Yes, I think so too
	AgreeExplain	Accepting contribution of another participant with elaboration	Yeah, it does. I think that if we are little treat about AI, it would help us to use the wall and therefore, bring the transparency bit
	NotAgreeSimple	Not accepting contribution of another participant without elaboration	I mean, I don't think there was a question there
	NotAgreeExplain	Not accepting contribution of another participant with elaboration	That sounds like an ideal situation. However, I again have to disagree with you about the um, it's turning into reality any points soon…

(Continues)

TABLE 1 (Continued)

Categories	Codes	Definitions	Examples
Socio-emotional interactions			
Socio-emotional interactions	PersonalFeel	Expressing personal emotions, subjective experiences or internal states, especially related to how they feel about a situation or task	I feel I am not creative enough and I am starting to depend on generative AI. I also like storytelling
	SocialEngage	Expressing social pleasantries, compliments, greetings or casual conversation, often without direct task-related interaction	Uh yes. Hi KAI. I am Alice [anonymised name]. Okay, thank you very much. I think that of today and of now thank you for being a good guy
	AI_SocialEngage	Mirroring or participating in social pleasantries, casual conversation or expressions of rapport-building	That's wonderful to hear! Take your time, I'm patient and here to listen…

Note: Apart from PersonalFeel, all other codes are also used to code GAI's responses which are marked with AI_[actualcode].

coding a second portion, bringing the total to approximately 17% (232 of 1317 utterances). At this phase, reliability, calculated prior to reconciliation, improved to $\kappa = 0.77$, indicating medium to high agreement, with no code below 0.68. Final codes were agreed upon for this portion. Rater 1 then codes the remaining dataset. To further assess the consistency of the coding, a third independent rater was subsequently recruited to blind double-code 25% of the full dataset. Agreement between Rater 1 and Rater 3 reached $\kappa = 0.79$ (0.67–0.85), supporting the reliability of the applied coding scheme and coded data used in subsequent analyses.

Sequential analysis

To uncover distinct types of cognitive and socio-emotional interactions in learners' engagements with GAI, temporal sequence analysis was employed to identify patterns by comparing the similarity of paired interactions. The analysis included 26 session sequences, comprising 1317 transcribed and coded meaningful discourse units. Each sequence represented a full interaction session between a student and KAI, capturing the temporal order of coded units, preserving the flow of adjacent transitions between utterances. These sequences included both learner and AI utterances, treating all interactions in a session as a unified temporal flow. For instance, transitions, such as 'learner's InfoSimple → learner's AskSimple' or 'GAI's AI_AskSimple → learner's AnsSimple' were considered valid. This approach ensured that all coded units, regardless of speaker or type, contributed to the analysis of interaction patterns. The Optimal Matching (OM) algorithm, implemented via the R packages *TraMineR* (Gabadinho et al., 2011) and *seqHMM* (Helske & Helske, 2017), was utilised to analyse interaction sequences. Each student sequence served as the unit of clustering, allowing participants to be directly mapped to cluster types and facilitating subsequent analyses, such as PMQ comparisons and ONA.

Based on the resulting distance matrix from OM, Ward's Clustering method was applied to hierarchically group sequences with similar patterns across sessions, using the *cluster* package in R (Maechler, 2018). The agglomerative coefficient, indicating the clustering strength, was 0.62. The optimal number of clusters was determined by evaluating the Silhouette coefficient for fit quality (see Table 2) and interpreting the dendrogram.

Ordered sequential analysis

Recent advancements in learning sciences and regulation research have integrated network analytics like epistemic network analysis (ENA) and sequential analysis to investigate process-oriented learning phenomena, including process mining and lag sequential analysis (Dang et al., 2024; Ouyang et al., 2023). Building on ENA, ONA extends its capabilities by modelling directed connections in data, incorporating event order to produce directed weighted networks instead of undirected models (Tan et al., 2023). Like ENA, ONA uses coded data to identify and measure connections, representing them in a metric space for statistical and visual comparisons (Shaffer et al., 2016; Shaffer & Ruis, 2017). However,

TABLE 2 AHC model fit statistic.

No. of cluster	2	3	4	5	6	7	8	9	10
Silhouette score	**0.130**	0.114	0.109	0.113	0.073	0.080	0.082	0.080	0.080

Note: The bolded value represents the highest silhouette score and was used to select the optimal number of clusters.

ONA's unique focus on event sequencing allows for the analysis of processes where the order of occurrences plays a crucial role, offering richer insights into learning dynamics by combining features of ENA and process mining.

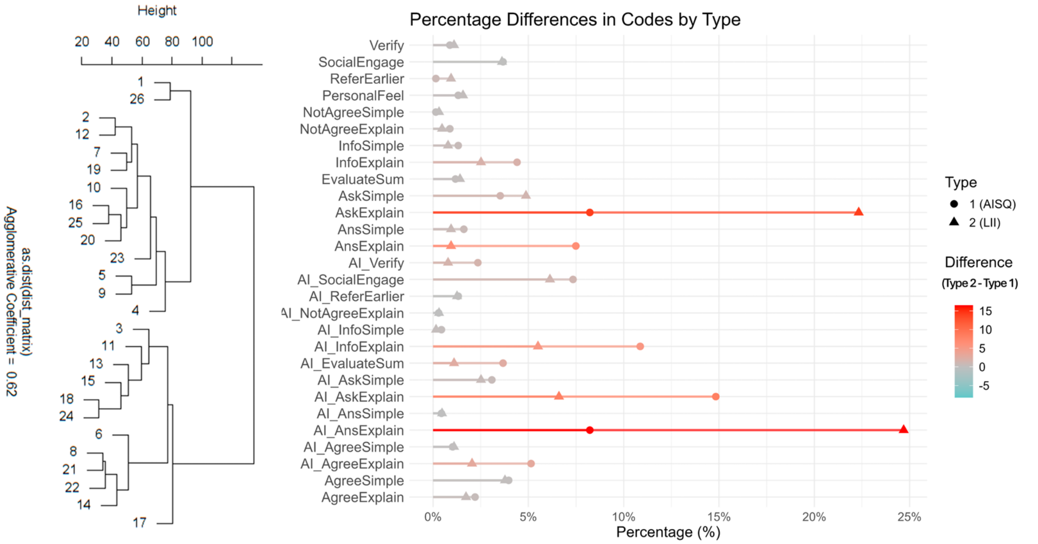
The key assumption of the method is that the structure of connections in the data is meaningful. This method is particularly relevant to our study, valuable for modelling the structural and sequential characteristics of cognitive and socio-emotional interactions in collaborative learning and knowledge construction (Tan et al., 2023). By capturing both the relationships and the directional co-occurrence of questions and answers, ONA effectively represents the dynamics of collaborative discussions between learners and GAI in an MR environment.

We defined the units of analysis as all lines of data associated with a single value of cluster subsetted by *sessionID* and *speakers*. Each ONA was constructed for a cluster and included all lines associated with a specific speaker (learner or AI) within a single learning session. The moving conversation window is four lines, decided based on the multiple discussion purposes that entail within each learner's or GAI's utterance. Directed networks were aggregated for each unit using binary summation, reflecting the presence or absence of co-occurrence between code pairs. The networks were normalised to account for variations in the number of coded lines across units before applying dimensional reduction. Directed network graphs were used for visualisation, where nodes represent codes, and edges indicate the direction and frequency of connections. Each unit of analysis was represented by (1) a plotted point in the low-dimensional space, corresponding to the network's location, and (2) a directed weighted network graph. Our model had co-registration correlations of 0.99 (Pearson) and 0.99 (Spearman) for the first dimension and co-registration correlations of 0.99 (Pearson) and 0.99 (Spearman) for the second. These measures indicate that there is a strong goodness of fit between the visualisation and the original model. A 3% minimum edge weight threshold was applied to reduce visual clutter while preserving essential information, a practice also seen in other studies, such as (Srivastava et al., 2022). Overlapping nodes were highlighted in orange to clarify potential ambiguities among codes.

RESULTS AND FINDINGS

RQ1: What are the key patterns of cognitive and socio-emotional interactions in human–AI collaborative learning an embodied GAI agent in MR?

This study examines learners' cognitive and socio-emotional interactions during the collaborative learning discussion through a multi-method analytical approach, incorporating QCA, sequential analysis and ONA. To address the first research question, Ward's hierarchical clustering, applied to 26 action sequences, identified two distinct collaborative patterns. Figure 2a illustrates the clustering of conversational interactions across the sessions. Based on the dendrogram and the optimal number of clusters determined using the Silhouette coefficient (see Table 2), the 26 input sessions were categorised into two primary types of interaction patterns. The first type contains the left 14 sessions, and the second type contains the right 12 sessions. A chi-squared test revealed a significant difference in the distribution of conversational actions between the two types ($\chi^2 = 200.56$, $df = 27$, $p < 0.01$), with a Cramér's V of 0.39 suggesting a moderate association. with a Cramér's V of 0.39 indicating a moderate association. This further highlights distinct variations in learners' cognitive and socio-emotional interactions with GAI between the two types. Figure 2b shows the frequency of cognitive and socio-emotional interaction



(a) **(b)**

FIGURE 2 (a) Hierarchical tree represents the result from Optimal Ward's clustering in which the number represents the session ID. (b) Frequency of codes by percentage in each type (Type 1 and Type 2).

exchanges across the two patterns. The results indicate comparable frequencies for most interaction types, particularly in socio-emotional dimensions, where both patterns show equivalent engagement. Socio-emotional interactions between learners and GAI account for approximately 10 to 12% of the overall interactions in both patterns. Earlier research has found that embodied agents in an interface produced more positive social interactions (Yee et al., 2007). In line with previous studies, our result recognised the emergence of socio-emotional interactions when learners interact with GAI in an embodied human-like interface and further examined its influence on learner-GAI dynamics and cognitive processes using ONA.

Meanwhile, Type 2 exhibits significantly higher frequencies of *AskExplain* and *AI_AnsExplain* interactions compared with Type 1, while Type 1 shows greater frequencies of *AnsExplain*, *AI_InfoExplain* and *AI_AskExplain* codes. These variations provide an initial indication that suggests differing interaction dynamics and potential shifts in the roles adopted by learners and GAI, with Type 1 interactions hinting at GAI taking a more active role in prompting and guiding discussions, and Type 2 interactions reflecting a more learner-driven inquiry approach. Therefore, we propose referring to Type 1 interactions initially as the AI-led Supported Exploratory Questioning (AISQ) type and Type 2 initially as Learner-Initiated Inquiry (LII) type.

Regarding the partner perceptions towards GAI, we report, for each type, the comparison of participants' responses to the baseline and post-survey done using paired Wilcoxon signed rank test. The test reveals a statistically significant increase from pre-test to post-test, indicating the experiment's positive effect (Table 3). For the AISQ group, the Wilcoxon test revealed a significant increase in scores from pre-test to post-test ($V=75$, $p=0.043$), with a mean difference of 0.68 (SD = 0.98). The effect size, measured by rank-biserial correlation, was 0.70, suggesting a substantial improvement. Similarly, for the LII type, the test also indicated a significant increase ($V=67$, $p=0.030$), with a mean difference of 0.72 (SD = 0.98) and an effect size of 0.71. These findings demonstrate that while both types experienced meaningful and comparable perception changes, the extent of improvement was slightly more pronounced in the LII type.

TABLE 3 Paired-sample Wilcoxon signed rank test of the Partner Model Questionnaire (Doyle et al., 2023).

Type		*N*	Mean	SD	*V*	*p*-Value	Rank-Biserial correlation
AISQ	Pre-	14	3.99	0.90	75	0.043	0.70
	Post-	14	4.67	0.77			
LII	Pre-	12	4.15	0.93	67	0.030	0.71
	Post-	12	4.88	0.91			

RQ2: How do the structural and temporal characteristics of key patterns in human–AI collaborative learning differ with respect to the cognitive and socio-emotional aspects?

To address RQ2 and, more critically, to gain deeper insights into the characteristics of each collaborative interaction pattern, we constructed and analysed the ONA graph (Figure 3) representing the cognitive and socio-emotional interactions within each type. This approach not only models the interactive, interdependent and temporal structures of each key pattern but also accounts for the order of events co-temporally, which can reveal additional insight into the clustered patterns that might remain elusive through quantitative metrics alone.

Figure 3a,b presents the individual unit networks for LII and AISQ, respectively, while Figure 3c illustrates the subtraction network between the two types, with differences represented by distinct colours. In these figures, squares denote centroids (ie, mean positions) for each type, and coloured dots represent individual learners and GAI. The node size indicates the frequency of a collaborative learning interaction that acts as a response to other actions. The more frequent that action was subsequent to other actions, the larger the corresponding node is. For example, in the LII type, *AskExplain* and *AI_AnsExplain* occurred as a response to other actions relatively more frequently, as also shown in Figure 3. This suggests that learners in the LII type often engage in explanatory questioning and elicit detailed responses from GAI, reflecting a pattern of interaction that emphasises inquiry and elaboration.

The subtraction networks in Figure 3c highlight the differences between the ordered networks of the two interaction patterns, shedding light on the distinct dynamics of learners engaging with GAI. The blue lines in the figure represent the strongest distinctions between the two types. For AISQ, the most dominant connections are between *AI_AskExplain*, *AnsExplain* and *AI_InfoExplain*, whereas for LII, the strongest connections are between *AskExplain* and *AI_AnsExplain*. From these connections alone—similar to insights gained through methods like ENA—a general interpretation might suggest that learners in the AISQ type demonstrate greater engagement and cooperation by elaborating on GAI's questions and potentially using GAI as a supplementary source of information. However, to move beyond surface-level insights and achieve a deeper understanding of these learning patterns, it is critical to examine the sequential structure of interactions, which the ONA technique effectively reveals.

As shown in Figure 3c, the connection from *AI_AnsExplain→AskExplain* emerges as a defining feature of LII interactions. This indicates a dynamic where learners, after receiving explanations from GAI, often follow up with additional questions or challenge its perspectives. This is further supported by observing that these two codes have connections and frequently succeeded learner's cognitive interactions, such as *ReferEarlier* or *InfoExplain*. This suggests that learners are engaging with the conversation, actively integrating new information with prior discussions and contributing their own perspective before asking GAI. Here,

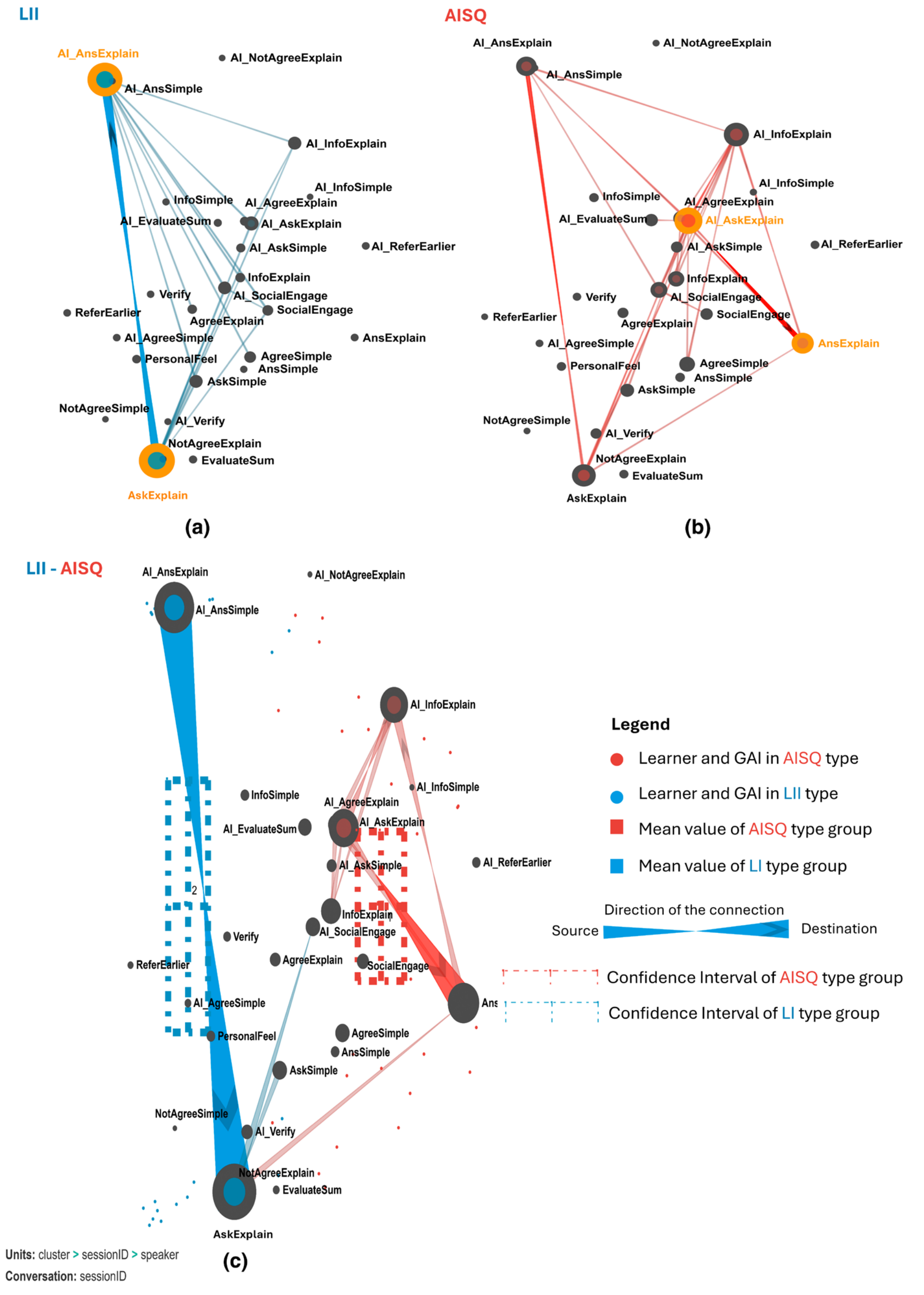


FIGURE 3 Ordered network analysis of learners in different types. (a) LII individual unit network. (b) AISQ individual unit network. (c) Subtraction network showing differences between groups.

GAI plays a responsive role, facilitating an inquiry-driven interaction that supports learners in exploring and broadening the discussion rather than plain information-providing. Meanwhile, AISQ interactions reflect a role where GAI acts as a guiding partner actively steering the interaction and learners engaging through detailed responses (*AI_AskExplain→AnsExplain* and *AnsExplain→AI_InfoExplain*). In this type, the presence of more connections involving preceding codes, such as *AI_EvaluateSum*, *AI_Verify* and *AI_ReferEarlier* suggests that GAI takes on a more reflective and confirmatory role.

We also performed a Mann–Whitney test to examine differences in the structure between two types of interaction patterns. Along the *x*-axis (dimension 1 after mean rotation), a Mann–Whitney test revealed a statistically significant difference between the AISQ type (Mdn = 0.24, N = 28) and the LII type (Mdn = −0.32, N = 24, U = 5.00, p = 0.00, r = 0.99). These results indicate that Type 1 and Type 2 generally demonstrate different patterns of cognitive interactions and socio-emotional interaction (along the *x*-axis).

While the differences in cognitive interaction structures between the two types are quite evident, what is particularly intriguing is what the ONA reveals about the socio-emotional dimensions. As observed, the level of socio-emotional interactions of both types is relatively comparable in terms of frequency as well as structure, evident in the connection of *SocialEngage* with other codes in both types. As previously discussed, the embodied form of GAI appears to facilitate more social pleasantries during interactions. A code of particular interest is *PersonalFeel*. In the LII type, there is a connection from *PersonalFeel→AskExplain*, suggesting learners tend to contextualise their inquiries through emotional expressions, making their questions more meaningful and relevant. In contrast, the connections involving *PersonalFeel* for AISQ type in the subtract network are not as prominent. This difference implies that, although socio-emotional interactions are present in both types, their integration with cognitive interactions varies. The expression of personal feelings in LII type may reflect a heightened interest in particular topics relevant to the learners, prompting them to probe more deeply and engage more thoroughly with the content.

DISCUSSION

This study aimed to explore the emerging phenomenon of human–AI collaboration and offer new insights into learning contexts involving embodied GAI agents within the evolving landscape of MR environments. Powered by advanced multimodal large language models (MLLMs), GAI enables highly human-like interactions, presenting novel opportunities and challenges for teaching and learning. While existing research has largely focused on the risks and ethical concerns surrounding these technologies in education (Dwivedi et al., 2023; Kasneci et al., 2023) there is growing recognition of the need to understand human–AI collaboration to harness GAI's potential for addressing evolving educational needs and societal demands (Cukurova, 2025; Järvelä et al., 2023).

Previous studies have highlighted the role of embodied GAI agents in enhancing learners' emotional experiences and recognising socio-emotional interactions, which in turn influence behaviour and engagement (Bickmore & Picard, 2005; Maldonado et al., 2017). However, these studies were conducted during a period when the capabilities of AI agents or embodied agents were relatively limited. Earlier models, while still valuable, lacked the sophisticated multimodal processing and flexibility capabilities now available, which are critical for adaptively interpreting and responding to socio-emotional cues in real time. For example, Adamson et al. (2014) demonstrated how simpler intelligent conversational agents could support collaboration with dynamic scaffolding prompts, albeit using less sophisticated architectures. On the contrary, recent efforts to apply learning analytics have yielded important insights into the processes of human–AI collaboration (eg, Cohn et al., 2024; Hong et al., 2021). They have, nonetheless, predominantly focused on text-based interactions, emphasising output-driven tasks and the final products of human–AI collaboration. Research involving GAI powered by MLLMs through other modalities, such as natural speech dialogue, remains scarce. This study aims to address this gap by exploring both the cognitive and socio-emotional aspects of interactions as well as the role of an embodied GAI agent within a MR context that enhances immersion and social realism. By doing so, we contribute to the understanding of human–AI collaboration in learning environments, specifically by offering insights into interaction patterns with GAI agents and how collaborative and regulatory mechanisms unfold within these GAI-integrated contexts.

Learning and learning regulation is inherently a social process (Vygotskii, 1978), where exchanged interactions and discussions play a pivotal role in constructing knowledge and deepening or forming shared understandings (Cohn et al., 2024; Dillenbourg et al., 2009). However, such interactions are not purely cognitive; they are profoundly shaped by socio-emotional dynamics (Isohätälä et al., 2017; Ouyang & Chang, 2019). Emotions, such as trust, curiosity (Vogl et al., 2019) and a sense of belonging not only encourage participation but also influence critical thinking and memory retention (Tyng et al., 2017). Positive emotional states can enhance learners' ability to process information deeply, think critically and sustain engagement, ultimately enriching the collaborative learning experience (Liu et al., 2023; So & Brush, 2008).

In line with this, the current investigation was able to identify two types of behaviour patterns employed by learners when collaborating with the GAI agent as a learning partner in the MR environment. These types exhibited significant variations in their interaction patterns and engagement levels at cognitive as well as socio-emotional aspects. The AISQ interactions type involves GAI acting as a guiding partner, often initiating prompts and elaborations to scaffold learners' understanding in a manner similar to a Socratic tutor (*AI_AskExplain→AnsExplain*). This form of interaction fosters structured cognitive engagement, as learners rely on the scaffolding and informational support provided by GAI to develop their understanding systematically. The LII interactions type, in contrast, involves a more inquiry-driven approach, where learners frequently initiate follow-up questions and challenge GAI's responses (*AI_AnsExplain→AskExplain*). These connections suggest that learners actively engage with GAI's responses, using them as a springboard for deeper exploration and critical thinking. This dynamic positions GAI as a responsive expert rather than a directive guide, reflecting a more exploratory and learner-led interaction. These findings align with the work of (De Smet et al., 2010; Jeremić et al., 2004), who observed similar dynamics in human–tutor settings, emphasising different characteristics as well as the value of structured support in the learning process.

A key finding of our study is the significant role of socio-emotional interactions in shaping cognitive processes and interaction dynamics. Both types of interaction—the AISQ type (guided, structured) and the LII type (inquiry-driven, exploratory)—exhibited consistent levels of socio-emotional interactions but manifested differently. In the AISQ interactions type, where GAI functions as a guiding partner, the weaker *PersonalFeel* connection did not diminish positive perceptions of the partnership. Indeed, structured guidance can effectively meet learners' needs even without overt socio-emotional cues, aligning with previous research on the benefits of a structured approach to personal tutoring (Watts, 2011). Conversely, the LII interactions type was inquiry-driven, marked by a strong transition from expressing personal emotions *PersonalFeel→AskExplain*. In fact, learners in the LII type displayed a greater improvement in their perception of GAI, KAI specifically, as a partner. Emotional exchanges seemed to create an initial connection, helping learners frame their questions in ways that were personally relevant and contextually meaningful. This approach seemed to allow learners to drive the interaction through their own interested inquiries, potentially enhancing the sense of learners' ownership. This illustrates how the socio-emotional context shapes how learners process information and engage with learning content (Immordino-Yang & Damasio, 2007; Zembylas, 2007).

The contrast between these two types of interactions indicates that while both socio-emotional interactions and scaffolding guidance are important, their effectiveness may depend on the type of interaction patterns, learner preferences and situational needs.

Structured and scaffolding guidance can satisfy learners' needs in a more directive, support-required context without explicit socio-emotional cues. On the other hand, socio-emotional interactions can enrich exploratory, self-directed learning by deepening personal engagement. Therefore, understanding the balance between these elements is crucial for designing AI learning partners that can adapt to various contexts and support diverse learner needs effectively.

This research presents several implications for educators, instructional designers and students within the education field. For educators, it uncovers specific cognitive and socio-emotional interaction patterns that emerge during collaborative learning with embodied GAI agents in MR contexts. These insights are crucial for designing targeted teaching strategies that enhance student skill development, focusing on leveraging structured support and socio-emotional interactions effectively. Instructional designers and educational technologists can leverage the identified functionalities of GAI, which include realistic gestures, voice modulation and natural movement, to create more effective educational tools that support both structured learning and authentic social interactions. For students, the study provides practical guidance on optimising their use of GAI, enabling them to effectively reflect and leverage GAI's roles—whether as a structured tutor providing scaffolding or as a collaborative peer facilitating inquiry-driven exploration.

LIMITATIONS AND FUTURE DIRECTIONS

Our study is not without limitations. The relatively small sample size and the specific MR context limit the generalisability of our findings. Future research should expand to include larger and more diverse participant groups to validate and extend the observed interaction patterns. For example, adopting a broader lens on the types of interactions, such as socio-cognitive or behavioural dimensions, might be a more comprehensive perspective on how learners engage and collaborate in MR environments. While it was beyond the focus of our study, future research can take into account aspects, such as knowledge gains or skill development to provide a more comprehensive understanding of GAI's influence on the collaborative learning process. The design does not include the incorporation of physical objects in the MR environment; this is another limitation, potentially constraining its overall effectiveness. Nonetheless, the study provides a naturalistic setting for dialogue-based learning and human-like interaction with KAI and virtual objects, while also sensing the real physical world, setting it apart from traditional VR or AR systems. Additionally, given the nature of MR and embodied agents, there are also promising avenues for integrating multimodal data with our analysis, such as gaze tracking, non-verbal cues and voice analysis, to provide deeper insights into the cognitive and socio-emotional interactions between learner and GAI.

CONCLUSION

This study enhances our understanding of the evolving role of generative AI in collaborative learning, highlighting the interplay between cognitive and socio-emotional dynamics in a ‘face-to-face’ collaboration with an embodied GAI within a MR environment. By examining distinct collaborative patterns, we provide valuable insights into how learners engage with AI, shedding light on the dynamics of human–AI collaboration in immersive, interactive settings, as well as the factors influencing the adoption and impact of educational technologies in this era. The versatility of GAI, particularly its ability to stimulate and respond to learners' socio-emotional interactions, positions it as a powerful tool not only for enhancing learning but also for fundamentally reshaping how humans learn, collaborate and regulate their learning processes. With technology continuously forging new pathways in education, including XR and the multiverse, so does the role of multimodal learning analytics becomes increasingly critical. It must advance to address the complexities of data collection, interpretation and analysis in these transformative learning systems.

ACKNOWLEDGEMENTS

This research has been funded by the Research Council of Finland (also known as Academy of Finland) (grants no.: 350249), and the University of Oulu profiling project Profi7 Hybrid Intelligence (352788). The study was conducted with the support of the LeaF Research Infrastructure at the University of Oulu. Open access publishing facilitated by Oulun yliopisto, as part of the Wiley - FinELib agreement.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy/ethical restrictions.

ETHICS STATEMENT

This research adheres to the ethical guidelines set forth by the Finnish National Board on Research Integrity TENK (2019) and the Ethics Committee of Human Sciences at the University of Oulu. This also complied with the European General Data Protection Regulation (GDPR) legislation. Informed consent was collected from all the participants.

ORCID

Belle Dang https://orcid.org/0009-0006-8734-6697
Luna Huynh https://orcid.org/0000-0001-5111-8277
Carolyn Rosé https://orcid.org/0000-0003-1128-5155
Sanna Järvelä https://orcid.org/0000-0001-6223-3668
Andy Nguyen https://orcid.org/0000-0002-0759-9656

REFERENCES

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, *24*(1), 92–124. https://doi.org/10.1007/s40593-013-0012-6

Bakhtiar, A., Webster, E. A., & Hadwin, A. F. (2018). Regulation and socio-emotional interactions in a positive and a negative group climate. *Metacognition and Learning*, *13*(1), 57–90. https://doi.org/10.1007/s11409-017-9178-x

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, *12*(2), 293–327.

Bower, M., Lee, M. J. W., & Dalgarno, B. (2017). Collaborative learning across physical and virtual worlds: Factors supporting and constraining learners in a blended reality environment. *British Journal of Educational Technology*, *48*(2), 407–430. https://doi.org/10.1111/bjet.12435

Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE Transactions on Man Machine Systems*, *11*(4), 190–202.

Chang, C.-Y., Kuo, H.-C., & Du, Z. (2023). The role of digital literacy in augmented, virtual, and mixed reality in popular science education: A review study and an educational framework development. *Virtual Reality*, *27*(3), 2461–2479. https://doi.org/10.1007/s10055-023-00817-9

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohn, C., Snyder, C., Fonteles, J. H., Ashwin, T. S., Montenegro, J., & Biswas, G. (2024). A multimodal approach to support teacher, researcher and AI collaboration in STEM +C learning environments. *British Journal of Educational Technology*, *56*(2), 595–620. https://doi.org/10.1111/bjet.13518

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, *32*(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, *56*, 469–488. https://doi.org/10.1111/bjet.13514

Dang, B., Nguyen, A., & Järvelä, S. (2024). The unspoken aspect of socially shared regulation in collaborative learning: AI-driven learning analytics unveiling 'silent pauses'. In *Proceedings of the 2024 Learning Analytics and Knowledge Conference (LAK24)*, Kyoto, Japan. https://doi.org/10.1145/3636555.3636874

De Smet, M., Van Keer, H., De Wever, B., & Valcke, M. (2010). Cross-age peer tutors in asynchronous discussion groups: Exploring the impact of three types of tutor training on patterns in tutor support and on tutor characteristics. *Computers & Education*, *54*(4), 1167–1181.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning: From design to orchestration. In N. Balacheff, S. Ludvigsen, T. Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning. Principles and products* (pp. 3–19). Springer.

Doyle, P. R., Gessinger, I., Edwards, J., Clark, L., Dumbleton, O., Garaialde, D., Rough, D., Bleakley, A., Branigan, H. P., & Cowan, B. R. (2023). *The partner modelling questionnaire: A validated self-report measure of perceptions toward machines as dialogue partners*. https://doi.org/10.48550/ARXIV.2308.07164

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Edwards, J., Nguyen, A., Lämsä, J., Sobocinski, M., Whitehead, R., Dang, B., Roberts, A., & Järvelä, S. (2025). Human-AI collaboration: Designing artificial agents to facilitate socially shared regulation among learners. *British Journal of Educational Technology*, *56*, 712–733. https://doi.org/10.1111/bjet.13535

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Flavián, C., Ibáñez-Sánchez, S., & Orús, C. (2019). The impact of virtual, augmented and mixed reality technologies on the customer experience. *Journal of Business Research*, *100*, 547–560. https://doi.org/10.1016/j.jbusres.2018.10.050

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, *40*, 1–37.

Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2024). The promise and challenges of generative AI in education. *Behaviour & Information Technology*, 1–27. https://doi.org/10.1080/0144929X.2024.2394886

Hamilton, A., Wiliam, D., & Hattie, J. (2023). The future of AI in education: 13 things we can do to minimize the damage. *EdArXiv*. https://doi.org/10.35542/osf.io/372vr

Helske, S., & Helske, J. (2017). Mixture hidden Markov models for sequence data: The seqHMM package in R. *arXiv Preprint arXiv:1704.00543*.

Hennessy, S., Cukurova, M., Lewin, C., Mavrikis, M., & Major, L. (2024). BJET editorial 2024: A call for research rigour. *British Journal of Educational Technology*, *55*(1), 5–9. https://doi.org/10.1111/bjet.13426

Higgins, E. T. (1999). Learning about what matters in the social world. *European Journal of Social Psychology*, *29*, 2–39.

Hodges, C. B., & Kirschner, P. A. (2024). Innovation of instructional design and assessment in the age of generative artificial intelligence. *TechTrends*, *68*(1), 195–199. https://doi.org/10.1007/s11528-023-00926-x

Holz, T., Campbell, A. G., O'Hare, G. M. P., Stafford, J. W., Martin, A., & Dragone, M. (2011). MiRA—Mixed reality agents. *International Journal of Human-Computer Studies*, *69*(4), 251–268. https://doi.org/10.1016/j.ijhcs.2010.10.001

Hong, A., Lunscher, N., Hu, T., Tsuboi, Y., Zhang, X., dos Reis, F., Alves, S., Nejat, G., & Benhabib, B. (2021). A multimodal emotional human–robot interaction architecture for social robots engaged in bidirectional communication. *IEEE Transactions on Cybernetics*, *51*(12), 5954–5968. https://doi.org/10.1109/TCYB.2020.2974688

Huang, X., & Lajoie, S. P. (2023). Social emotional interaction in collaborative learning: Why it matters and how can we measure it? *Social Sciences & Humanities Open*, *7*(1), 100447. https://doi.org/10.1016/j.ssaho.2023.100447

Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain, and Education*, *1*(1), 3–10. https://doi.org/10.1111/j.1751-228X.2007.00004.x

Isohätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research*, *81*, 11–24. https://doi.org/10.1016/j.ijer.2016.10.006

Isohätälä, J., Näykki, P., & Järvelä, S. (2020). Cognitive and socio-emotional interaction in collaborative learning: Exploring fluctuations in students' participation. *Scandinavian Journal of Educational Research*, *64*(6), 831–851. https://doi.org/10.1080/00313831.2019.1623310

Järvelä, S., & Hadwin, A. (2024). Triggers for self-regulated learning: A conceptual framework for advancing multimodal research about SRL. *Learning and Individual Differences*, *115*, 102526. https://doi.org/10.1016/j.lindif.2024.102526

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2021). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*, *72*, 101203. https://doi.org/10.1016/j.learninstruc.2019.04.004

Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, *54*(5), 1057–1076. https://doi.org/10.1111/bjet.13325

Järvenoja, H., & Järvelä, S. (2009). Emotion control in collaborative learning situations: Do students regulate emotions evoked by social challenges? *British Journal of Educational Psychology*, *79*(3), 463–481. https://doi.org/10.1348/000709909X402811

Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, *206*, 104898. https://doi.org/10.1016/j.compedu.2023.104898

Jeong, S., Clyburn, J., Bhatia, N. S., McCourt, J., & Lemons, P. P. (2022). Student thinking in the professional development of college biology instructors: An analysis through the lens of sociocultural theory. *CBE—Life Sciences Education*, *21*(2), ar30. https://doi.org/10.1187/cbe.21-01-0003

Jeremić, Z., Devedžić, V., & Gašević, D. (2004). An Intelligent Tutoring System for learning design patterns. In *Proceedings of the ICWE2004 Workshop on Adaptive Hypermedia and Collaborative Web-Based Systems, Munich, Germany*.

Kara, M., & Çakıcı Alp, N. (2024). Assessing the adoption of the Yavuz battleship application in the mixed reality environment using the technology acceptance model. *Multimedia Systems*, *30*(2), 76. https://doi.org/10.1007/s00530-024-01277-0

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kauffeld, S., Lehmann-Willenbrock, N., & Meinecke, A. L. (2018). The advanced interaction analysis for teams (act4teams) coding scheme. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The Cambridge handbook of group interaction analysis* (1st ed., pp. 422–431). Cambridge University Press. https://www.cambridge.org/core/product/identifier/9781316286302%23CN-bp-21/type/book_part

Kim, J., Merrill, K., Jr., Xu, K., & Kelly, S. (2022). Perceived credibility of an AI instructor in online education: The role of social presence and voice features. *Computers in Human Behavior*, *136*, 107383. https://doi.org/10.1016/j.chb.2022.107383

Kuo, Y.-C., Chu, H.-C., & Huang, C.-H. (2015). A learning style-based grouping collaborative learning approach to improve EFL students' performance in English courses. *Journal of Educational Technology & Society*, *18*(2), 284–298.

Kwon, K., Liu, Y.-H., & Johnson, L. P. (2014). Group regulation and social-emotional interactions observed in computer supported collaborative learning: Comparison between good vs. poor collaborators. *Computers & Education*, *78*, 185–200. https://doi.org/10.1016/j.compedu.2014.06.004

Lee, G.-. G., Mun, S., Shin, M.-. K., & Zhai, X. (2025). Collaborative learning with artificial intelligence speakers: Pre-service elementary science teachers' responses to the prototype. *Science & Education*, *34*(2), 847–875. https://doi.org/10.1007/s11191-024-00526-y

Lee, W., & Reeve, J. (2020). Brain gray matter correlates of general psychological need satisfaction: A voxel-based morphometry study. *Motivation and Emotion*, *44*(1), 151–158. https://doi.org/10.1007/s11031-019-09799-1

Li, M., Guo, F., Wang, X., Chen, J., & Ham, J. (2023). Effects of robot gaze and voice human-likeness on users' subjective perception, visual attention, and cerebral activity in voice conversations. *Computers in Human Behavior*, *141*, 107645. https://doi.org/10.1016/j.chb.2022.107645

Liaw, S. Y., Tan, J. Z., Lim, S., Zhou, W., Yap, J., Ratan, R., Ooi, S. L., Wong, S. J., Seah, B., & Chua, W. L. (2023). Artificial intelligence in virtual reality simulation for interprofessional communication training: Mixed method study. *Nurse Education Today*, *122*, 105718. https://doi.org/10.1016/j.nedt.2023.105718

Liu, Z., Yu, P., Liu, J., Pi, Z., & Cui, W. (2023). How do students' self-regulation skills affect learning satisfaction and continuous intention within desktop-based virtual reality? A structural equation modelling approach. *British Journal of Educational Technology*, *54*(3), 667–685. https://doi.org/10.1111/bjet.13278

Maas, M. J., & Hughes, J. M. (2020). Virtual, augmented and mixed reality in K–12 education: A review of the literature. *Technology, Pedagogy and Education*, *29*(2), 231–249. https://doi.org/10.1080/1475939X.2020.1737210

Maechler, M. (2018). *Cluster: Cluster analysis basics and extensions*. R Package Version 2.0. 7–1.

Maldonado, H., Lee, J.-E. R., Brave, S., Nass, C., Nakajima, H., Yamada, R., Iwamura, K., & Morishima, Y. (2017). We learn better together: Enhancing elearning with emotional characters. In *Computer supported collaborative learning 2005* (pp. 408–417). Routledge.

Molenaar, I. (2022). The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. *Computers and Education: Artificial Intelligence*, *3*, 100070. https://doi.org/10.1016/j.caeai.2022.100070

Naik, A., Yin, J. R., Kamath, A., Ma, Q., Wu, S. T., Murray, R. C., Bogart, C., Sakr, M., & Rosé, C. P. (2025). Providing tailored reflection instructions in collaborative learning using large language models. *British Journal of Educational Technology*, *56*(2), 531–550. https://doi.org/10.1111/bjet.13548

Näykki, P., Isohätälä, J., & Järvelä, S. (2021). "You really brought all your feelings out"—Scaffolding students to identify the socio-emotional and socio-cognitive challenges in collaborative learning. *Learning, Culture and Social Interaction*, *30*, 100536. https://doi.org/10.1016/j.lcsi.2021.100536

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, *49*, 847–864. https://doi.org/10.1080/03075079.2024.2323593

Nguyen, A., Kremantzis, M., Essien, A., Petrounias, I., & Hosseini, S. (2024). Editorial: Enhancing student engagement through artificial intelligence (AI): Understanding the basics, opportunities, and challenges. *Journal of University Teaching & Learning Practice*, *21*(6). https://doi.org/10.53761/caraaq92

Ouyang, F., & Chang, Y. (2019). The relationships between social participatory roles and cognitive engagement levels in online discussions. *British Journal of Educational Technology*, *50*(3), 1396–1414. https://doi.org/10.1111/bjet.12647

Ouyang, F., Xu, W., & Cukurova, M. (2023). An artificial intelligence-driven learning analytics method to examine the collaborative problem-solving process from the complex adaptive systems perspective. *International Journal of Computer-Supported Collaborative Learning*, *18*(1), 39–66. https://doi.org/10.1007/s11412-023-09387-z

Owoseni, A., Kolade, O., & Egbetokun, A. (2024). Enhancing personalised learning and student engagement using generative AI. In A. Owoseni, O. Kolade, & A. Egbetokun (Eds.), *Generative AI in higher education* (pp. 123–150). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-60179-8_5

Sandhaus, H., Gu, Q., Parreira, M. T., & Ju, W. (2024). Student reflections on self-initiated GenAI use in HCI education. *arXiv:2410.14048. arXiv*. https://doi.org/10.48550/arXiv.2410.14048

Shaffer, D., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, *3*(3), 9–45. https://doi.org/10.18608/jla.2016.33.3

Shaffer, D., & Ruis, A. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In *Handbook of learning analytics*. Society for Learning Analytics Research (SoLAR).

Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). Exploring students' knowledge construction strategies in computer-supported collaborative learning discussions using sequential analysis. *Journal of Educational Technology & Society*, *17*(4), 216–228.

So, H.-J., & Brush, T. A. (2008). Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education*, *51*(1), 318–336. https://doi.org/10.1016/j.compedu.2007.05.009

Srivastava, N., Fan, Y., Rakovic, M., Singh, S., Jovanovic, J., Van Der Graaf, J., Lim, L., Surendrannair, S., Kilgour, J., Molenaar, I., Bannert, M., Moore, J., & Gasevic, D. (2022). Effects of internal and external conditions on strategies of self-regulated learning: A learning analytics study. In *LAK22: 12th International Learning Analytics and Knowledge Conference*, Newport Beach, CA (pp. 392–403). https://doi.org/10.1145/3506860.3506972

Tan, Y., Ruis, A. R., Marquart, C., Cai, Z., Knowles, M. A., & Shaffer, D. W. (2023). Ordered network analysis. In C. Damşa & A. Barany (Eds.), *Advances in quantitative ethnography* (Vol. *1785*, pp. 101–116). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-31726-2_8

Tyng, C. M., Amin, H. U., Saad, M. N., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, *8*, 235933.

Vogl, E., Pekrun, R., Murayama, K., Loderer, K., & Schubert, S. (2019). Surprise, curiosity, and confusion promote knowledge exploration: Evidence for robust effects of epistemic emotions. *Frontiers in Psychology*, *10*, 2474.

Vygotskii, L. S. (Lev S., 1896–1934). (1978). *Mind in society: The development of higher psychological processes / L. S. Vygotsky; edited by Michael Cole … [Et al.]*. Harvard University Press.

Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2024). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*, *37*(4), 814–840. https://doi.org/10.1080/09588221.2022.2056203

Watts, T. E. (2011). Supporting undergraduate nursing students through structured personal tutoring: Some reflections. *Nurse Education Today*, *31*(2), 214–218. https://doi.org/10.1016/j.nedt.2010.06.005

Webb, N. M., & Mastergeorge, A. M. (2003). The development of students' helping behavior and learning in peer-directed small groups. *Cognition and Instruction*, *21*(4), 361–428. https://doi.org/10.1207/s1532690xci2104_2

Weber, F., Wambsganss, T., & Soellner, M. (2023). Design and evaluation of an AI-based learning system to foster students' structural and persuasive writing in law courses. *ICIS 2023* Proceedings, 9. https://aisel.aisnet.org/icis2023/learnandiscurricula/learnandiscurricula/9

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In *Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum Associates Publishers.

Wolf, E., Fiedler, M. L., Döllinger, N., Wienrich, C., & Latoschik, M. E. (2022). Exploring presence, avatar embodiment, and body perception with a holographic augmented reality mirror. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Christchurch, New Zealand (pp. 350–359). https://doi.org/10.1109/VR51125.2022.00054

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, *21*(1), 40. https://doi.org/10.1186/s41239-024-00468-z

Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, *34*(3), 327–343. https://doi.org/10.1017/S0958344022000039

Yee, N., Bailenson, J. N., & Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Montréal, Canada (pp. 1–10). https://doi.org/10.1145/1240624.1240626

Zamecnik, A., Kovanović, V., Joksimović, S., Grossmann, G., Ladjal, D., & Pardo, A. (2024). The perceptions of task cohesion in collaborative learning teams. *International Journal of Computer-Supported Collaborative Learning*, *19*(3), 369–393. https://doi.org/10.1007/s11412-024-09424-5

Zembylas, M. (2007). Emotional capital and education: Theoretical insights from Bourdieu. *British Journal of Educational Studies*, *55*(4), 443–463. https://doi.org/10.1111/j.1467-8527.2007.00390.x

How to cite this article: Dang, B., Huynh, L., Gul, F., Rosé, C., Järvelä, S., & Nguyen, A. (2025). Human–AI collaborative learning in mixed reality: Examining the cognitive and socio-emotional interactions. *British Journal of Educational Technology*, *00*, 1–24. https://doi.org/10.1111/bjet.13607

APPENDIX A

THE LEARNING ENVIRONMENT AND EMBODIED AGENT

The mixed reality learning environment, developed in Unity for use with Meta Quest 3 headsets, features KAI, an embodied AI agent created with the Inworld AI platform (https://inworld.ai/), integrated into Unity using the Inworld API. KAI's ability to process and interact multimodally with human-like behaviour is enabled by a suite of machine learning models. These include large language models (LLMs), text-to-speech (TTS), automatic speech recognition (ASR), emotion engines, memory and machine vision, etc. KAI's cognition was engineered by researchers to align with the study's pedagogical objectives. This included tailoring its knowledge base by leveraging the LLM's general knowledge and domain-specific content on AI ethics, curated by researchers with expertise in the topic. During the generation response, KAI also considers the student's conversation history for personalised interactions. To support critical thinking, KAI was designed to have the tendency to emulate Socratic teaching strategies, prompting learners with open-ended, probing questions that encourage elaboration and reflection. This was not based on fixed scripts but rather modelled as a dialogue partner with an inquisitive disposition, intended to simulate a thoughtful, sometimes challenging peer or mentor. KAI's design incorporates human-like traits, such as specific personality, dialogue style and minor flaws, such as being overly inquisitive, to create an engaging and relatable learning partner. The agent was iteratively tested by researchers and refined to ensure functionality, knowledge accuracy, discussion flow and alignment with dialogue-based learning's goal for higher education students.

The learning task, involving free-flow conversations with KAI simulating expert-guided dialogue, was designed by learning science researchers. This aimed to facilitate critical thinking and ethical reasoning for higher education students. Through open-ended discussions, learners were encouraged to articulate their opinions and reflect on ethical dilemmas in AI. Supporting this were interactive UI panels, designed using Unity's Interaction Toolkit, Mixed Reality Toolkit and Grabbable UI Canvas Tablet (https://assetstore.unity.com/), which presented videos, flashcards, instructions and transition aids. These panels structured the learning session while allowing flexibility for exploration. This setup combined immersive interaction with scaffolded support, enabling learners to develop a deeper understanding through dialogue and active engagement with the content.