Journal of Computer Assisted Learning

WILEY

Journal of Computer Assisted Learning

ORIGINAL ARTICLE

Boosting Student Engagement in STEM: Integrating Large Language Model-Based Virtual Agents Into Alternate Reality Games

Minkai Wang[1] | Jingdong Zhu[1] | Gwo-Jen Hwang[2,3,4] | Shao-Chen Chang[5] | Qi-Fan Yang[6] | Di Zhang[7]

[1]College of Education, Zhejiang University of Technology, Hangzhou, China | [2]Graduate Institute of Educational Information and Measurement, National Taichung University of Education, Taichung, Taiwan | [3]Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei, Taiwan | [4]College of Management, Yuan Ze University, Taoyuan, Taiwan | [5]Department of International Bachelor Program in Informatics and the Department of Information Communication, Yuan Ze University, Taoyuan, Taiwan | [6]College of Education, Fujian Normal University, Fuzhou, China | [7]Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua, China

Correspondence: Di Zhang (zhangdi0909@126.com)

Received: 29 March 2025 | **Revised:** 3 September 2025 | **Accepted:** 2 October 2025

Funding: This work was supported by the National Social Science Fund of China, CCA230341.

Keywords: game-based learning | large language models | learning engagement | STEM | virtual agent

ABSTRACT

Background: STEM education aims to develop innovation and problem-solving skills through interdisciplinary learning, yet struggles to foster student engagement and interdisciplinary thinking. Whilst alternate reality games (ARGs) can boost motivation via game-based problem-solving, integrating large language models (LLMs) remains underexplored. LLM-based virtual agents offer new opportunities for adaptive support.

Objectives: This study aimed to investigate the effectiveness of an LLM-assisted ARG system (LLM-ARG) in enhancing academic performance, metacognitive awareness, and engagement.

Methods: A quasi-experimental study compared LLM-ARG with conventional ARG methods amongst primary school students. The experimental group used LLM-ARG with personalised virtual agent support, whilst the control group employed a conventional ARG with a traditional, rule-based virtual agent that offered only pre-scripted feedback. Data were collected through pre- and post-tests, metacognitive awareness questionnaires, and interaction logs. ANCOVA and correlation analyses were conducted.

Results and Conclusions: LLM-ARG significantly improved learning achievements and metacognitive awareness compared to conventional ARG. High-frequency interactions promoted exploration but did not consistently enhance problem-solving, whilst low-frequency interactions led to higher success via goal-directed strategies. Metacognitive competence emerged as a key predictor of academic performance, highlighting the need to balance exploration with efficiency. This study demonstrates how LLM-driven scaffolding supports diverse learning strategies and promotes adaptive learning in STEM education.

1 | Introduction

Science, technology, engineering, and mathematics (STEM) education is widely recognised as a crucial approach for enhancing students' scientific literacy, critical thinking, and abilities to solve interdisciplinary problems (English 2023). Integrating science, technology, engineering, and mathematics into real-world scenarios, STEM education fosters students' autonomous exploration and innovative practises, deepening their disciplinary knowledge and honing their problem-solving skills

Summary

- What is currently known about this topic
  - STEM education struggles with fostering student engagement and interdisciplinary thinking.
  - Alternate reality games (ARGs) can boost motivation via real-world problem-solving.
  - Large language models (LLMs) offer adaptive support but are rarely used in ARGs.
- What does this paper add
  - Introduces an LLM-assisted ARG system to improve STEM engagement and outcomes.
  - Demonstrates that personalised virtual agents enhance metacognitive awareness.
  - Identifies a balance between exploration and efficiency as key to effective learning.
- Implications for practise and/or policy
  - Applying LLM-driven ARGs can significantly enhance personalised STEM education.
  - Teachers should balance exploratory learning with efficiency-driven strategies.
  - Policymakers should promote AI tools that build students' metacognitive and strategic learning skills.

(Tan et al. 2023). As educational technologies evolve, the importance of STEM education has become more prominent in the educational sector (Li et al. 2020). Yet, its practical application encounters numerous challenges, such as the abstract nature of scientific knowledge (Qian and Choi 2023), the complexity of learning materials (Gao et al. 2020), and traditional teaching methods falling short when it comes to motivating students intrinsically, thus limiting their learning motivation and effectiveness (Lin and Wang 2021). From a cognitive science perspective, these challenges can impose a high cognitive load on students, hindering their ability to effectively process information and engage in deep learning (Sweller 1994). The question of how to effectively create realistic learning environments that enable students to comprehend and apply abstract and complex concepts is now a central concern in educational research and practise (Chernikova et al. 2020).

Alternate Reality Games (ARGs) are interactive narratives that use the real world as a platform, often blurring the line between the game and reality (Örnebring 2007). Unlike typical games that are confined to a screen, ARGs integrate digital content with real-world locations and tasks to create an immersive, trans-media experience (Jagoda et al. 2015). Grounded in situated learning theory, they are extensively employed to kindle students' passion for learning and to promote their active engagement (Gilliam et al. 2017; Meekaew and Ketpichainarong 2021). Through the creation of puzzles and challenges, ARGs effectively boost students' motivation to learn (Hou et al. 2023; Sofianidis et al. 2024) and facilitate their problem-solving in blended virtual and real settings (Stylianidou et al. 2020). Although ARGs are widely recognised for their significant potential to foster active inquiry, research indicates that many students struggle with the organisation and processing of complex real-world information, which impacts their performance in ARG tasks (Whitton et al. 2014). For example, the overload of multimodal information, including visual, auditory, and textual data, can increase cognitive load, thereby reducing students' efficiency of information processing (Moon and Ryu 2021). To unlock the full potential of ARGs in STEM education, it is vital to provide tailored learning support that assists students in navigating challenges encountered during the inquiry process, thus boosting their inquiry skills and educational outcomes (Liang et al. 2021).

Chatbots driven by generative artificial intelligence, acting as virtual agents, have shown substantial promise in advancing human-computer interactions within educational settings (Yan et al. 2024). Large Language Models (LLMs) employed in these chatbots have witnessed rapid advancements recently, particularly in the fields of knowledge retrieval, text recognition, and multimodal interactions, which have notably increased the chatbots' potential in educational applications (Kuhail et al. 2023). LLM-based chatbots, by providing instantaneous feedback and supplementary information, effectively aid students in problem solving, and are recognised as personal tutors or learning assistants that can be seamlessly incorporated into educational systems (Dokukina and Gumanova 2020; Essel et al. 2022). Supported by extensive data, these chatbots not only manage information efficiently, but also enhance student engagement and problem resolution, delivering tailored learning feedback (Chen et al. 2023; Kasneci et al. 2023). Nonetheless, the integration of LLM technology into complex situational learning tasks remains underexplored, with most studies focusing primarily on its application in online learning platforms (Ooi et al. 2023). Whilst prior research has established the benefits of adaptive feedback, the unique affordances of LLMs—such as their ability to engage in natural language dialogue, provide explanatory feedback, and offer metacognitive scaffolding—remain underexplored within immersive learning environments (Du 2025; El Hajji et al. 2025).

To address these deficits, this study developed an LLM-assisted ARG system (LLM-ARG) that leverages a personalised virtual agent to provide adaptive scaffolding. From a theoretical standpoint, such an agent can support learning in two critical ways: (1) by reducing extraneous cognitive load through just-in-time information and feedback (Sweller 1994), and (2) by fostering self-regulated learning (SRL) skills, such as planning, monitoring, and reflection, which are crucial for developing metacognitive awareness (Zimmerman 1995). Building on this theoretical framework, this study investigates the effectiveness of the LLM-ARG system by specifically examining its impact on students' (1) learning achievements, (2) metacognitive awareness (as a key outcome of SRL scaffolding), and (3) behavioural engagement within the game environment. Accordingly, the pivotal research questions are:

1. Do students learning with LLM-ARG achieve better learning performance and metacognitive awareness than those learning with conventional ARG (C-ARG)?
2. What are the differences in the behavioural engagement and interaction patterns of students using LLM-ARG and those using C-ARG?
3. What are the relationships amongst interaction frequency, metacognitive skills, and academic performance?

2 | Literature Review

2.1 | Alternate Reality Games

To foster enthusiasm for STEM learning amongst students, ARGs effectively bolster the interactions between digital game tasks and real-world settings (Hou et al. 2023). Serving as a game-based learning paradigm, ARGs engage students through interactive narratives and challenges, thereby igniting their spirit of active inquiry (Vayanou et al. 2019) and increasingly becoming a focal point in educational research (Elsom et al. 2023; Hou et al. 2023). By incorporating digital tools and problem-solving techniques, ARGs craft an immersive learning environment that not only enhances the realism of the educational experience, but also inspires students to actively engage and investigate (Kassutto et al. 2021).

ARGs have been successfully integrated into classroom educational practises (Christopoulos et al. 2023; Jesionkowska et al. 2020), with studies affirming their ability to markedly boost student engagement, ignite motivation, and enhance the overall learning experience (Srisawasdi and Panjaburee 2019). By forging connections between learners and real-world scenarios, ARGs employ gamified narratives and comprehensive feedback mechanisms to guide students in refining their inquiry strategies (De Beer and Bothma 2016; Habibi et al. 2022). Research indicates that ARGs foster advanced cognitive skills such as critical thinking, problem-solving, and teamwork through interdisciplinary tasks (Gilliam et al. 2017; Tulloch et al. 2021). In STEM education, ARGs serve as an ideal platform that enables students to hone their critical thinking and collaborative skills within interdisciplinary settings (Stylianidou et al. 2020). These complex, multidisciplinary tasks require students to amalgamate knowledge from various domains to tackle dynamic educational challenges. Such environments not only aid in the transfer and practical application of knowledge, but also potentially enhance students' abilities in deep thinking and innovation. This potential benefit merits further investigation in the context of STEM education.

Despite their potential, students encounter numerous challenges during the learning process in ARGs. The inherent uncertainty of puzzles in ARGs, coupled with the complexities of real-world scenarios, heightens the difficulty of problem solving, often impeding students' ability to collect clues and synthesise information for task completion (Gilliam et al. 2017). Throughout the ARG tasks, students may feel frustrated, discouraged, and distracted, which can prevent them from achieving the desired learning outcomes. Some may lose motivation altogether if the challenges surpass their abilities (Richey et al. 2019). Given the complexity and diversity of these tasks, ineffective organisation and internalisation of information can disrupt students' cognitive construction processes (Glazewski and Ertmer 2020). Furthermore, an overload of tasks or information can lead to cognitive overload, undermining the efficiency and effectiveness of their knowledge construction (Sewell et al. 2020; Xu and Ouyang 2022). Successfully navigating these challenges requires learners to possess strong metacognitive skills to monitor their understanding, manage cognitive load, and adjust their problem-solving strategies (Seufert 2018). However, students often struggle to self-regulate effectively in such complex environments (Jin et al. 2023). This highlights a critical need for adaptive, personalised feedback within ARGs to serve as an external scaffold, thereby fostering the metacognitive awareness necessary for successful inquiry-based learning.

2.2 | Virtual Agent

Virtual agents facilitate learner autonomy by establishing roles akin to tutors or partners, offering essential guidance and feedback throughout the learning journey (Seo et al. 2021). In educational settings, these agents assume a supportive role by fostering independent learning amongst students, without overtly directing the learning trajectory (Wang et al. 2024). Learning is inherently a dynamic process that evolves not only through personal intellectual effort, but also via interactions with others or with technological systems (Pettersson 2021). Virtual agents adeptly mimic the functions of teachers or peers, providing timely feedback and emotional support when challenges arise, thus enhancing the social dimensions of learning. Additionally, they enrich opportunities for collaboration and communication, enabling deeper knowledge construction (Chen et al. 2020; Song et al. 2019). By transitioning traditional face-to-face engagements into novel forms of human-machine collaboration, virtual agents effectively enrich the interactive experience across varied learning settings (Guo et al. 2024).

With virtual agents' support, students can effectively complete learning tasks and practises, benefiting from timely feedback that assists them in identifying and rectifying errors (Oker et al. 2020; Sikström et al. 2022). Additionally, virtual agents boost students' self-efficacy and learning motivation through emotional feedback, which also strengthens their social presence in the learning process, reduces anxiety, and ignites their internal drive to learn (Chou et al. 2022; Fountoukidou et al. 2019). This emotional support not only increases students' feelings of security and belonging in virtual environments, but also motivates them to engage more actively in learning activities (Schlimbach et al. 2024). Research indicates that virtual agents not only spur learning motivation but also enhance the efficiency with which students execute tasks, thereby improving their performance and cognitive growth in complex learning scenarios (Krassmann et al. 2019; Lin et al. 2020). Virtual agents are emerging as potent educational tools that provide sustained support, boosting engagement and a sense of achievement amongst students tackling intricate tasks (Sinatra et al. 2021). Whilst the benefits demonstrated in these studies are significant, the effectiveness of most traditional virtual agents is constrained by their underlying technology. Typically reliant on pre-scripted dialogue trees and rule-based systems, their ability to provide truly adaptive, explanatory feedback or understand novel student inquiries is limited (Han et al. 2025). This rigidity can hinder the development of sophisticated learning processes that require more than pre-programmed hints (Al-Abri 2025). Therefore, to enable the kind of dynamic, conversational scaffolding needed for effective learning in varied STEM educational settings, this research turns to the advanced capabilities of LLMs.

2.3 | Personalised Learning Based on LLM

Personalised learning refers to tailoring instructional content, pace, and approach to meet the unique needs and abilities of individual learners (Zhang et al. 2022). Whilst traditional adaptive systems often personalise by recommending pre-defined content pathways (Kasneci et al. 2023), the unique affordance of an LLM is its ability to provide dynamic, conversational, and individualised support in real-time. An LLM is a natural language processing technology trained on extensive datasets to effectively understand and generate human language (Thirunavukarasu et al. 2023), with applications ranging from content summarisation to automated question-answering (Kasneci et al. 2023; Yang et al. 2024). As LLM capabilities advance, they are increasingly used to provide this kind of personalised, data-driven support (Chang et al. 2024).

Functioning as virtual agents, LLM-powered chatbots offer robust learning support by moving beyond simple feedback to enhance the learning experience through engagement and personalization (Chen et al. 2024; Lin and Mubarok 2021). By analysing students' real-time progress, LLMs can generate tailored recommendations or prompts to help them overcome challenges (Hu et al. 2024) and can accommodate diverse proficiency levels through dynamic adjustments, thus enabling adaptive learning (Goslen et al. 2024). Whilst prior research has shown these systems can improve general outcomes like learning achievements and motivation (Hew et al. 2023; Lee et al. 2022), their specific capacity to deliver timely, explanatory feedback is particularly well-suited for fostering self-regulated learning skills (Bibauw et al. 2019; Xia et al. 2023). Therefore, this study chooses to focus specifically on how this LLM-driven personalised scaffolding impacts metacognitive awareness and behavioural engagement, two key components of self-regulated learning.

Integrating learning support into game environments has been shown to significantly enhance learners' acquisition of knowledge and skills in game-based settings (Liao et al. 2019). However, if the learning support is poorly integrated, it may lead to increased cognitive load, disrupting students' gaming experience (Bainbridge et al. 2022). Virtual agents offer a promising solution by seamlessly embedding learning support into games, enhancing both learning outcomes and the overall gaming experience. This approach achieves the dual objectives of sparking interest and promoting education (Guo et al. 2024). Accordingly, this study incorporated LLM virtual agents into ARG, leveraging the immersive nature of ARG and the interactivity of virtual agents to investigate their potential effects on students' STEM learning performance, particularly in terms of knowledge acquisition, skill development, and engagement.

3 | ARG Design With an LLM-Assisted Virtual Agent Approach

3.1 | System Framework

As shown in Figure 1, the LLM-ARG system was designed to facilitate students' engagement with STEM curricula in a simulated real-world environment. Central to the system is 'V,' an LLM-Assisted Personal Virtual Agent, which supports students in conducting active inquiries within STEM. The system comprises three core modules: the Game module, the Learning module, and the PVA module (Personal Virtual Agent module). The Game module manages the gaming aspects, including the game plot, quests, non-player characters (NPCs), and exploration scenes. The Learning module oversees the educational components embedded within the game, such as learning rules, materials, tests, and feedback mechanisms. The PVA module

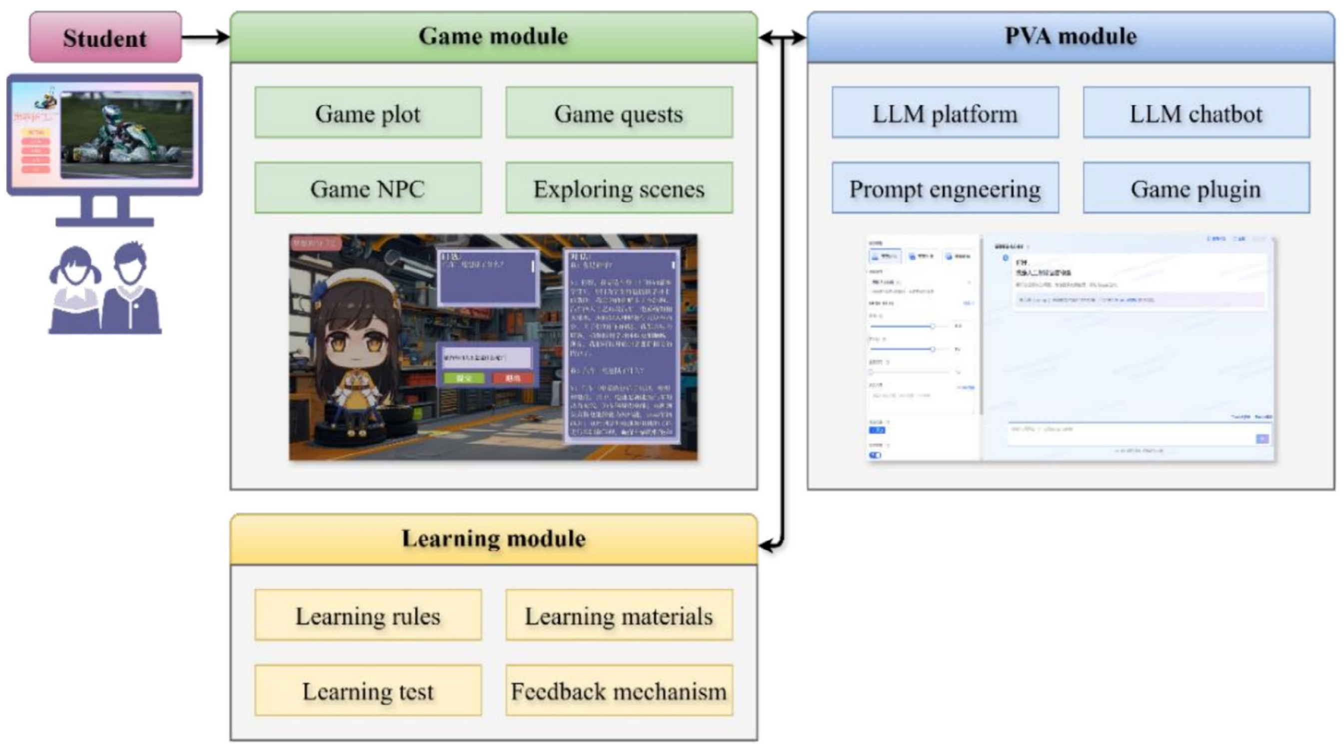


FIGURE 1 | LLM-ARG system framework.

integrates the LLM functionality, enabling seamless interaction with both the Game and Learning modules through components such as the LLM platform, chatbot, prompt engineering, and game plugin. This cohesive framework ensures a dynamic and immersive learning experience for students.

Figure 2 illustrates the ARG system workflow, where students engage in learning and complete all ARG tasks with the assistance and support of the virtual agent, 'V.' The system includes four game tasks: 'Accepting the Invitation,' 'Kart Structure,' 'Three Electrical Systems,' and 'Four Key Processes,' each mapped to a planned learning pathway for students. After completing one task, students can proceed to the next or choose to review previous content. Upon entering a task, students explore the scene to learn relevant knowledge points and collect knowledge cards through their exploration. They are also required to complete test questions to achieve the task objectives. Additionally, students can review their acquired knowledge cards and game achievements in the Library and Achievement Gallery sections on the home page, further reinforcing their learning outcomes.

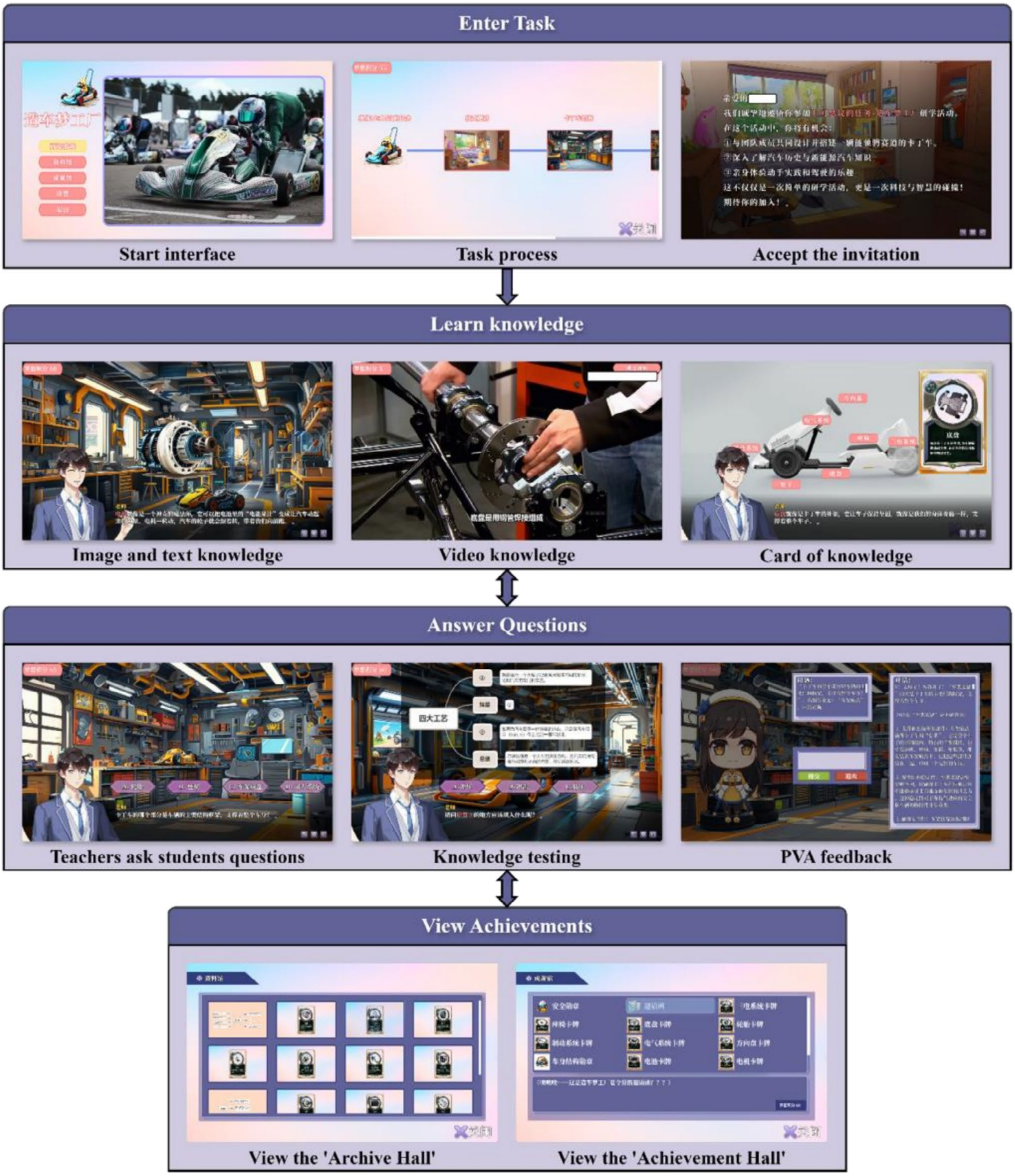


FIGURE 2 | The system flow of the ARG system.

3.2 | STEM Learning Content

STEM inquiry activities were designed around the four dimensions of science, technology, engineering, and mathematics, with clearly defined learning content and objectives that aimed to help students build a holistic knowledge framework and develop multifaceted skills through interdisciplinary tasks, as shown in Figure 3. These inquiry activities were introduced through game task modules, specifically 'Accepting the Invitation,' 'Kart Structure,' 'Three Electrical Systems,' and 'Four Key Processes.' Crucially, these game modules were designed to mirror authentic real-world processes, requiring students to bridge digital gameplay with real-world scientific and engineering knowledge. Each module provides students with an initial understanding of the relevant learning content and foundational skills. By participating in interactive gameplay, students can engage with key knowledge points and receive immediate feedback, serving as a preparatory step for deeper exploration and learning in subsequent tasks.

In the science dimension, students learn fundamental methods of scientific inquiry, such as conducting knowledge searches and organising information, whilst applying logical reasoning to analyse specific challenges, including the load distribution of the kart frame and the energy transmission pathways of the battery. The technology dimension focuses on data collection, tool operation, and environmental interaction skills, encouraging students to deepen their technical understanding through data visualisation and practical applications. In the engineering dimension, students evaluate and optimise the stability of the kart frame, honing their system optimization skills through prototype design, iterative testing, and performance refinement. Finally, the mathematics dimension emphasises the development of quantitative analysis skills by engaging students in tasks such as measuring component dimensions, calculating speed-to-weight ratios, and applying geometric principles to enhance kart performance.

3.3 | Personalised Virtual Agent Strategy

In designing the functionalities of the LLM virtual agent 'V', we adopted a learner-centred approach that provides three primary types of adaptive feedback, which are used consistently throughout this paper: (1) Answer Feedback, which diagnosed conceptual errors and delivered tailored explanations after students submitted responses; (2) Task-Oriented Feedback, which activated milestone reminders and reflective prompts if learners deviated from goals; and (3) Achievement Feedback, which triggered in-game rewards and progress summaries upon subgoal completion. This strategy ensures that feedback dynamically adapts to gameplay contexts whilst preserving pedagogical coherence. Table 1 illustrates these mechanisms with concrete examples within the 'Kart Structure' task.

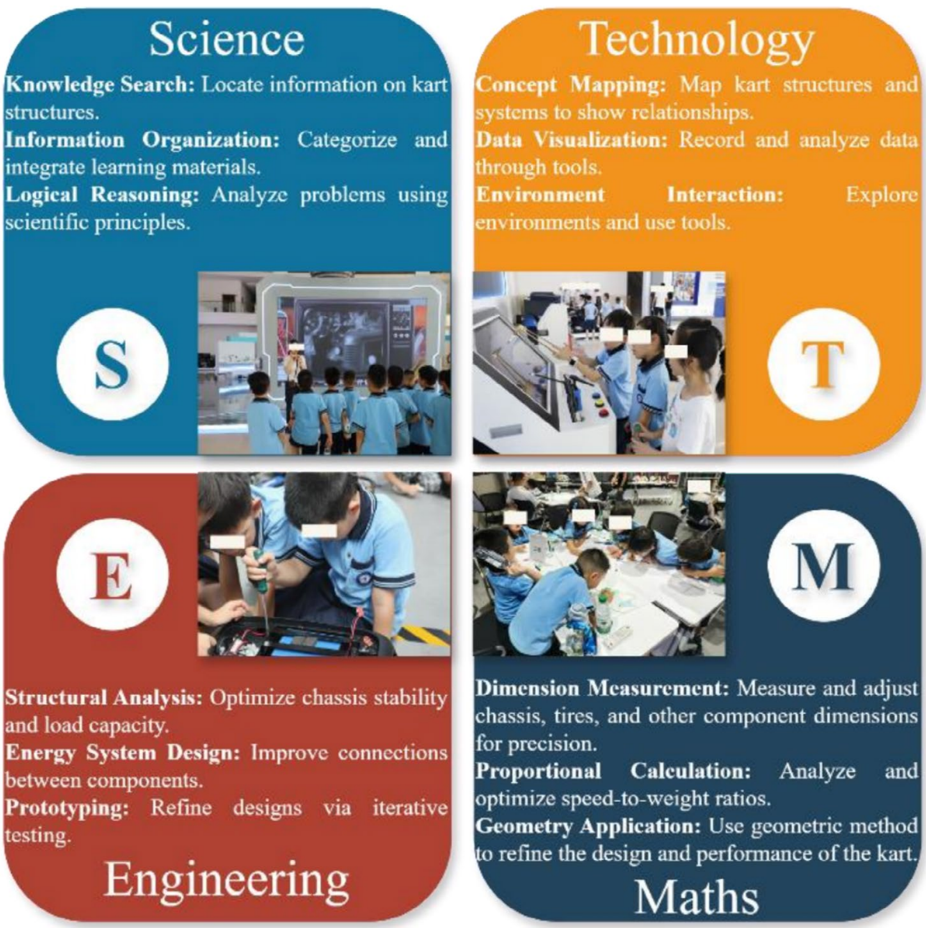


FIGURE 3 | STEM learning content.

TABLE 1 | LLM-driven feedback mechanisms in the 'kart structure' task.

Gameplay loop	Design purpose	LLM interaction method (prompt strategy + feedback type)	Concrete case
Task interaction	1. Reinforce conceptual accuracy 2. Cultivate inquiry-driven learning attitudes	Prompt: 'If answer = correct > 1. Praise 2. Explain chassis mechanics 3. Encourage inquiry'	'Great job! The chassis is indeed the main structural framework of the kart. Think about why it supports the entire body.'
	1. Diagnose knowledge gaps 2. Guide self-correction strategies	Prompt: 'Detect error_type = conceptual_misalignment > 1. Normalise mistake 2. Clarify tire role 3. Scaffold hints'	'It seems you are focusing on the tire, but remember that the chassis serves as the backbone of the structure.'
	1. Prevent frustration 2. Promote real-world connection	Prompt: 'If repeated_errors = 2 > 1. Highlight key terms 2. Suggest observational tasks 3. Trigger reflection'	'Try observing how the chassis supports the car body. It might help you rethink the structural framework.'
	1. Address persistent misconceptions 2. Train metacognitive monitoring	Prompt: 'If repeated_errors ≥ 3 > 1. Activate metacognitive questioning 2. Contrast functions 3. Suggest self-monitoring'	'Let's pause and consider: What is the primary purpose of the chassis compared to the power system?'
Objective tracking	1. Align actions with learning goals 2. Bridge current and future knowledge	Prompt: 'Post-task > 1. Extract key concepts 2. Generate personalised plan 3. Preview next topic'	'You've mastered the chassis concept! Next, we'll explore how the power system integrates with it. Stay curious!'
Reward unlocking	1. Enhance intrinsic motivation 2. Foster knowledge transfer through gamification	Prompt: 'Upon reward_unlock > 1. Contextual praise 2. Link reward to skills 3. Inspire creative application'	'Congratulations on unlocking the motor card! Think about how the chassis and motor work together in real applications.'

The system interface, shown in Figure 4, serves as the central hub for these interactions. All feedback from agent 'V' is delivered through the Dialogue Window. The process is initiated when a student submits a response in the Question Window, which automatically triggers Answer Feedback. Following this, the student can choose to engage further by asking follow-up questions. This entire interaction sequence—from receiving the initial automatic feedback to any subsequent, student-chosen dialogue—is defined as engaging with feedback. This clarifies that whilst initial feedback is automatically provided, deeper engagement is optional, allowing students to tailor the level of support they receive.

The LLM-ARG system employs this structured three-phase adaptive support framework to enhance learning effectiveness, engagement, and self-regulation. In the task execution phase, 'V' delivers real-time Answer Feedback by diagnosing errors and offering targeted conceptual clarifications. As students progress, the objective tracking phase utilises Task-Oriented Feedback to help students monitor their learning trajectories, evaluate performance, and generate personalised learning plans. Finally, the achievement reinforcement phase leverages Achievement Feedback through gamified mechanisms to enhance engagement by visualising progress and acknowledging learning milestones. Through these interconnected phases, the LLM-ARG system dynamically adapts to diverse learning needs, balancing exploratory refinement with structured goal attainment to optimise both cognitive development and sustained engagement.

3.4 | System Description

The overall technical framework of this study is illustrated in Figure 5, which is structured as a cyclic interaction model composed of four key modules: Students Interacting with Agent 'V', Data Processing and Transmission, Platform Interaction with Baidu AI Cloud Qianfan, and Feedback and Optimization. This framework emphasises the dynamic and continuous interaction between the virtual agent and the learning environment, leveraging the advanced capabilities of the Baidu AI Cloud Qianfan Platform as the technical backbone. The Qianfan platform is an all-in-one service platform launched by Baidu, offering a range of large model services, including ERNIE-Bot, as well as access to third-party large models. It is equipped with a comprehensive toolchain for the development and application of large models. The integration of ERNIE-Bot-4 into the LLM-ARG system allows for adaptive and intelligent interaction, significantly enhancing the quality and effectiveness of automated responses.

The LLM-ARG system is composed of four core modules. The first module is Students Interacting with Agent 'V'. In this module,

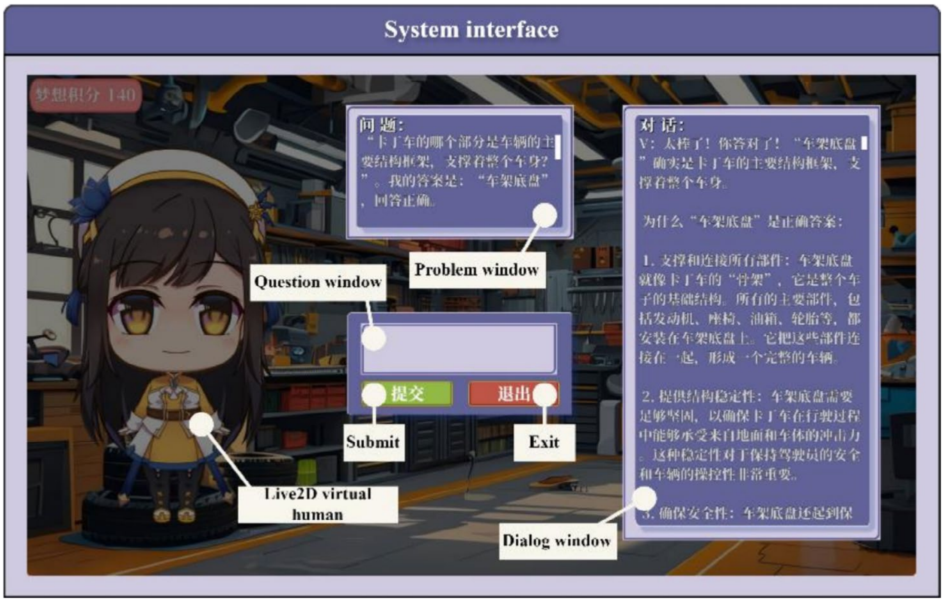


FIGURE 4 | System interface.

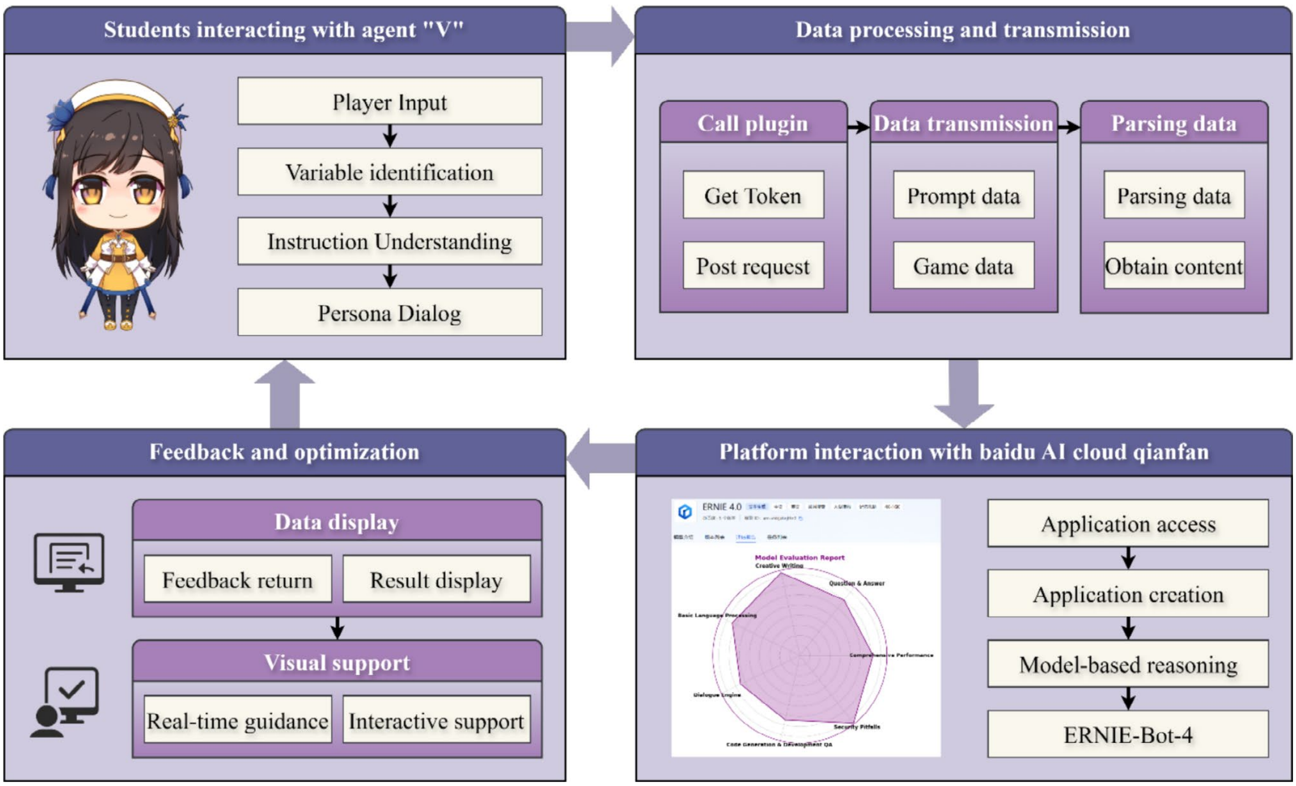


FIGURE 5 | Technology roadmap.

students interact with the virtual agent through various inputs, including player input, variable identification, instruction understanding, and persona dialogue. Agent 'V' dynamically interprets and responds to player actions and queries, delivering context-aware feedback that supports problem-solving and learning. The persona dialogue feature allows for personalised guidance, making the interaction experience more engaging and informative.

The second module is Data Processing and Transmission, which handles data interaction between the game environment and the Baidu AI Cloud Qianfan Platform. It consists of three key components: Call Plugin, Data Transmission, and Parsing Data Plugin. The Call Plugin manages obtaining Access Tokens and submitting HTTP requests to the Qianfan platform for model invocation. The Data Transmission component processes prompt data and game data to facilitate real-time interaction between the game engine and the large model. The Parsing Data Plugin is responsible for extracting and processing dialogue content from the model output, ensuring that the feedback provided by 'V' is relevant and coherent.

The third module is Platform Interaction with Baidu AI Cloud Qianfan, where the ERNIE-Bot-4 model, hosted on the Qianfan platform, conducts model-based reasoning and dynamically generates context-aware responses. The integration with the Qianfan platform ensures smooth and effective communication between the game and the model, leveraging Qianfan's capabilities for efficient large model utilisation.

The fourth module is Feedback and Optimization, which forms a critical part of the closed-loop interaction. After receiving responses from the model, the system enters the feedback loop. This module consists of Data Display and Visual Support. In Data Display, the system presents feedback return and result display to the game engine, providing users with clear and concise information. In Visual Support, the system provides real-time guidance and interactive support within the game interface, allowing students to receive instant feedback and make prompt adjustments. This feedback loop facilitates continuous interaction and re-engagement, allowing learners to refine their actions and enhance their problem-solving strategies.

The cyclic structure of the system ensures that after feedback is processed and presented, students can resume interaction with Agent 'V', forming a closed-loop learning cycle. This dynamic process helps maintain engagement whilst allowing continuous improvement through iterative feedback and response adaptation. By seamlessly integrating game-based interaction with model-driven reasoning, the system offers a coherent and efficient educational experience.

4 | Experimental Design

4.1 | Participants

To evaluate the learning effectiveness of this approach, a quasi-experimental design was adopted, involving 57 students aged 12–13 from two classes at a primary school in eastern China. These students had over 2 years of information technology education and possessed basic computer skills, enabling them to engage in digital game-based learning. One class was designated as the experimental group and the other as the control group. The experimental group, consisting of 28 students (14 boys and 14 girls), utilised the LLM-ARG approach, whilst the control group, comprising 29 students (15 boys and 14 girls), employed the C-ARG approach. All students in both groups were taught by the same teacher to ensure consistency in instruction.

The experimental group received real-time feedback through the LLM-ARG system, including knowledge-based feedback, task-oriented feedback, and achievement-driven motivational feedback. Knowledge-based feedback was designed to reinforce learned content, task-oriented feedback supported the construction of cognitive structures, and achievement-based feedback motivated students by rewarding their progress with achievements and items. In contrast, the control group's C-ARG system employed a traditional rule-based e-learning approach. The agent's feedback was limited to pre-scripted responses triggered by predefined rules (e.g., confirming a correct answer or providing a generic hint after an incorrect one). Crucially, this system lacked the LLM's capacity for dynamic, conversational interaction; it could not interpret student-initiated questions in natural language, diagnose the underlying nature of a misconception, or provide personalised metacognitive prompts. This design allows for a direct comparison between a standard game-based learning environment with basic feedback and one enhanced with novel LLM-driven, dialogic scaffolding, which is the focus of our study.

4.2 | Experimental Procedure

To evaluate the effectiveness of the LLM-ARG system, we utilised the interdisciplinary STEM curriculum 'Dream Car Factory' as the experimental teaching content. The experimental procedure is illustrated in Figure 6. Over the course of 5 weeks, learners participated in a structured STEM learning process designed to integrate ARG-based instructional interventions.

In Week 1, learners engaged in introductory activities, which began with a 20-min session to introduce learning activities and rules. Following this, students completed a 30-min pretest to assess academic performance and metacognitive awareness. These introductory activities ensured that all participants had a clear understanding of the learning objectives and initial baseline data before engaging in the main instructional content. During Weeks 2 to 4, both the experimental group and the control group participated in core instructional activities designed to build STEM knowledge and skills through lectures and hands-on problem-solving tasks. Following the instructional activities, each group completed a 60-min ARG task independently. The LLM-ARG group interacted with the personalised virtual agent 'V', which provided adaptive, real-time support, whilst the C-ARG group followed the same instructional activities using a system that provided only traditional, non-adaptive guidance, the specifics of which are detailed in the Participants section. Both groups were exposed to identical scenarios, educational content, and task durations to ensure consistency in the learning environment. In Week 5, a 60-min posttest was conducted to evaluate academic performance and post-intervention metacognitive awareness. This design ensured that the ARG intervention was systematically integrated into the

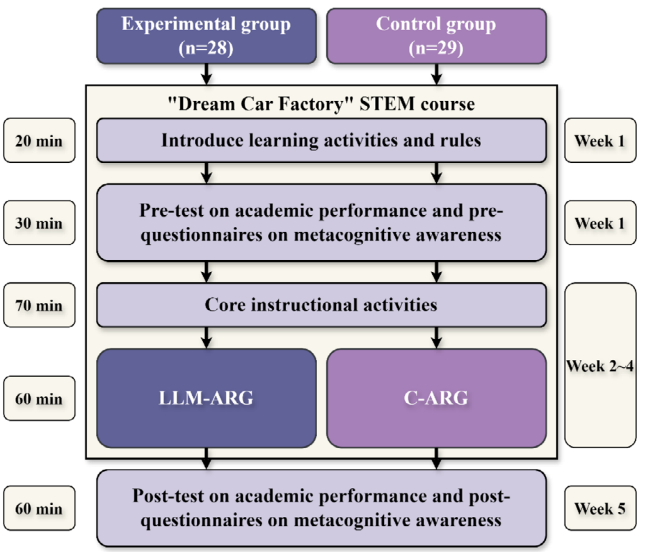


FIGURE 6 | Experimental procedure.

standard curriculum, maintaining consistent instructional time and task structure for both groups.

4.3 | Data Collection

The measurement tools used in this study included tests and questionnaires designed to assess learning achievements and metacognitive awareness. All pre- and posttests were rigorously reviewed by teachers and experts to ensure that participants could understand the questions accurately.

At the initial stage of the study, a pretest was conducted to ascertain the prior knowledge of students in both the experimental and control groups before they engaged in the learning activities. In contrast, the posttests were intended to examine differences in the learning achievements of the two groups after employing different learning modalities. Both tests covered knowledge related to kart structure, three electrical systems, and the four key processes, each consisting of 20 multiple-choice questions with a total score of 100 points. All tests were designed by the research team and were meticulously reviewed by experts and teachers with over 10 years of teaching experience to ensure the validity of the assessments. The KR-20 values before and after the tests were 0.72 and 0.74, respectively, indicating acceptable internal consistency of the tests.

Metacognitive awareness refers to an individual's ability to reflect on and adjust their own cognitive processes. The metacognitive awareness questionnaire was based on the measurement approach by Schraw and Dennison (1994) and revised by Lai and Hwang (2015). This questionnaire consisted of five items, including questions such as, 'When I complete a task, I ask myself if what I have learned from the task is sufficient.' It used a 5-point Likert scale for scoring (1 = *strongly disagree*; 5 = *strongly agree*), and with a Cronbach's alpha of 0.83, it demonstrated good reliability.

To analyse learning behaviours, student interactions within the ARG environment were captured in detailed system logs. This study adopted the behavioural analysis framework proposed by Liang et al. (2024), resulting in the coding system shown in Table 2. The appropriateness and clarity of this coding scheme were validated through discussions with two experienced teachers prior to the study. Crucially, the data logging and coding process was fully automated by the game system to ensure objectivity and eliminate the need for manual coding. The system logged specific events as follows: 'E' (Exploring) was recorded when a student navigated to a new game scene; 'R' (Reading learning materials) was recorded when a student opened a knowledge card and kept it on screen for more than 3 s; and 'S' (Skipping learning materials) was recorded if the card was closed in less than that time. All other codes (e.g., A, C, I, F, Q) were also logged automatically based on discrete user actions. Because this process was automated, inter-rater reliability analysis was not applicable. This automated process ensured a reliable and detailed record of the gaming experience, which was saved for subsequent analysis.

4.4 | Data Analysis

All quantitative data were analysed using statistical software. To examine the effectiveness of the LLM-ARG system, two separate analyses of covariance (ANCOVA) were conducted to compare the posttest scores of learning achievement and metacognitive awareness between the experimental and control groups, with their respective pretest scores serving as the covariate. To investigate the differences in behavioural patterns, a lag sequential analysis was performed on the coded interaction logs from the system. The significance of behavioural transitions was determined by calculating z-scores from the adjusted residuals, with a z-score greater than 1.96 considered statistically significant ($p < 0.05$). Finally, to explore the relationships amongst interaction frequency, metacognitive skills, and academic performance

TABLE 2 | Coding schema of game behaviour.

Code	Code name	Student learning behaviours in the game
E	Exploring	Student explores learning scenarios
R	Reading learning materials	Student reads graphic and video learning materials
S	Skipping learning materials	Student skips reading graphic and video learning materials
A	Answering question	Student answers a question in the game
C	Answering the question correctly	Student answers a question in the game correctly
I	Answering the question incorrectly	Student answers a question in the game incorrectly
I1	Answering the question incorrectly for the first time	Student answers a question in the game incorrectly for the first time
I2	Answering the question incorrectly for the second time	Student answers a question in the game incorrectly for the second time
I3	Answering the question incorrectly for the third time	Student answers a question in the game incorrectly for the third time
F	Choosing to receive feedback from V	Student actively selects to receive feedback from V, including answer feedback, task feedback, and prop feedback.
Q	Q&A with V	Student interacts with V and asks V questions

within the experimental group, a Pearson correlation analysis was conducted.

5 | Experimental Results

5.1 | Learning Achievement

This study employed Analysis of Covariance (ANCOVA) to investigate differences in the learning achievements of the experimental and control groups. The instructional method was set as the independent variable, with posttest scores as the dependent variable and pretest scores as the covariate. The Levene's test for equality of variances did not reach statistical significance ($F=0.580$, $p=0.450$), confirming the assumption of homogeneity of variances. The test for homogeneity of regression slopes indicated that there was no significant interaction between groups concerning the pretest scores ($p=0.927$), verifying that the linear relationships met the fundamental assumptions of ANCOVA. Table 3 presents the descriptive statistics for both groups. After controlling for pretest scores, ANCOVA revealed a significant difference in the posttest scores between the two groups ($F=4.258$, $p=0.044$). According to Ferguson (2009), a partial eta squared (η^2) of 0.073 indicates a small effect size. This suggests that students using the LLM-ARG significantly outperformed those using C-ARG, demonstrating that the LLM-based virtual agent ARG enhanced students' learning achievements.

5.2 | Metacognitive Awareness

This study employed ANCOVA to investigate the differences in the metacognitive awareness scores of the experimental and control groups. The instructional method was designated as the independent variable, with post-questionnaire metacognitive awareness scores serving as the dependent variable and pre-questionnaire scores as the covariate. The Levene's test for equality of variances did not reach statistical significance ($F=1.932$, $p=0.170$), confirming the assumption of homogeneity of variances. The test for homogeneity of regression coefficients indicated no significant interaction between the groups and the pretest scores ($p=0.554$), demonstrating that the data met the fundamental assumptions of ANCOVA. Table 4 presents the descriptive statistics for both groups. After accounting for the pretest scores, ANCOVA revealed a significant difference in the scores of the two groups ($F=4.480$, $p=0.039$). According to Ferguson (2009), a partial η^2 of 0.077 indicates a small effect size. This suggests that students using the LLM-ARG significantly outperformed those using C-ARG in terms of metacognitive awareness, indicating that the LLM-based virtual agent ARG enhanced students' metacognitive awareness.

5.3 | Lag Sequential Analysis

Lag sequential analysis, a statistical method for analysing the dependency of event sequences in time-series data, was employed to reveal patterns and trends in interactive behaviour sequences. Students in the experimental group generated a total of 2306 behaviour codes, whilst those in the control group produced 1745. Subsequently, the frequencies of various behaviour sequences were calculated, and an adjusted residuals table was produced using matrix methods. The residuals table consisted of rows representing the initial student behaviours and columns corresponding to subsequent behaviours. Adjusted residuals were computed through matrix operations, and sequences were considered statistically significant ($p<0.05$) when Z-values exceeded 1.96.

Tables 5 and 6 display the results of the adjusted residuals analysis, whilst Figures 7 and 8 illustrate the pathways of learning behaviours, highlighting significant differences between the experimental and control groups. In these diagrams, arrows indicate the direction of behaviour transitions, the thickness of the lines represents the significance level, and the Z-values next to the arrows denote statistical significance. Black lines represent behaviour sequences common to both groups, whilst red lines indicate significant behaviour sequences unique to the experimental group. Different types of behaviour codes are marked with corresponding coloured blocks.

5.4 | Comparison of Behaviour Sequences Between Experimental and Control Groups

The significant behaviour sequences identified in the experimental group included E→E, E→R, R→S, R→A, S→A, A→C, A→I1, C→F, I1→F, I2→F, I3→F, F→Q, Q→E, Q→R, Q→C,

TABLE 3 | The ANCOVA results of learning achievement.

Group	N	Mean	SD	Adjusted mean	SE	F	η^2
Experimental group	28	69.29	7.90	69.29	1.01	4.258*	0.073
Control group	29	66.38	6.39	66.38	0.99		

*$p<0.05$.

TABLE 4 | The ANCOVA results of metacognitive awareness.

Group	N	Mean	SD	Adjusted mean	SE	F	η^2
Experimental group	28	4.15	0.34	4.15	0.04	4.480*	0.077
Control group	29	4.05	0.27	4.05	0.04		

*$p<0.05$.

TABLE 5 | Results of residuals transition in the experimental group.

	E	R	S	A	C	I1	I2	I3	F	Q
E	9.69*	29.4*	−1.19	−5.91	−5.91	−4.2	−1.69	−0.97	−7.98	−9.49
R	−5.44	−4.89	9.2*	43.42*	−6.5	−4.62	−1.86	−1.07	−8.77	−10.44
S	−1.12	−0.44	−0.26	8.53*	−1.28	−0.91	−0.37	−0.21	−1.73	−2.06
A	−5.53	−6.5	−1.28	−6.35	18.04*	33.17*	−1.82	0.05	−8.57	−10.19
C	−5.53	−6.5	−1.28	−6.35	−6.35	−4.51	−1.82	−1.04	34.6*	−10.19
I1	−3.93	−4.62	−0.91	−4.51	−4.51	−3.2	−1.29	−0.74	24.58*	−7.24
I2	−1.58	−1.86	−0.37	−1.82	−1.82	−1.29	−0.52	−0.3	9.9*	−2.92
I3	−0.91	−1.07	−0.21	−1.04	−1.04	−0.74	−0.3	−0.17	5.69*	−1.68
F	−7.47	−8.45	−1.73	−8.57	−8.57	−6.09	1.66	−0.52	−11.57	37.92*
Q	15.42*	3.7*	−1.99	−9.88	11.67*	−7.02	4.78*	3.3*	−13.33	0.44

*$p < 0.05$.

TABLE 6 | Results of residuals transition in the control group.

	E	R	S	A	C	I
E	14.37*	18.18*	−2.89	−10.22	−10.22	−11.49
R	−10.13	−8.86	11.37*	35.52*	−8.86	−9.95
S	−2.87	−2.5	−0.71	11.37*	−2.5	−2.81
A	−10.13	−8.86	−2.5	−8.86	7.78*	20.29*
C	17.93*	9.51*	−2.48	−8.77	−8.77	−9.85
I	−11.39	−9.95	−2.81	−9.95	20.29*	12.04*

*$p < 0.05$.

Q → I2, and Q → I3. In contrast, the control group's significant behaviour sequences included E → E, E → R, R → S, R → A, S → A, A → C, A → I, C → E, C → R, I → C, and I → I.

Both groups exhibited similarities in exploration, reading materials, and answering questions. They tended to continue exploring after exploration (E → E) and chose to read materials after exploring (E → R). However, they often skipped parts of the content whilst reading (R → S), indicating that students might selectively skip based on interest or time management.

Additionally, after reading materials, students attempted to answer questions (R → A) but frequently gave incorrect answers. Both groups showed a tendency for incorrect answers to outnumber correct ones, as reflected in the experimental group's A → I1, A → I2, A → I3 and the control group's A → I transitions, indicating challenges in applying learned knowledge. These patterns suggest that regardless of virtual agent support, students exhibited similar learning behaviours in self-directed tasks, especially in exploring and problem-solving.

In the experimental group, students frequently interacted with the virtual agent following both correct and incorrect responses (C → F, I1 → F, I2 → F, I3 → F), engaging in repeated interactions (F → Q, Q → Q). These interactions led to higher correct answer rates (Q → C) and reduced error rates (Q → I2, Q → I3). This continuous feedback loop not only improved accuracy but also stimulated motivation, encouraging students to explore further (Q → E) and read materials (Q → R). These findings highlight the important role of the virtual agent in promoting both learning motivation and outcomes.

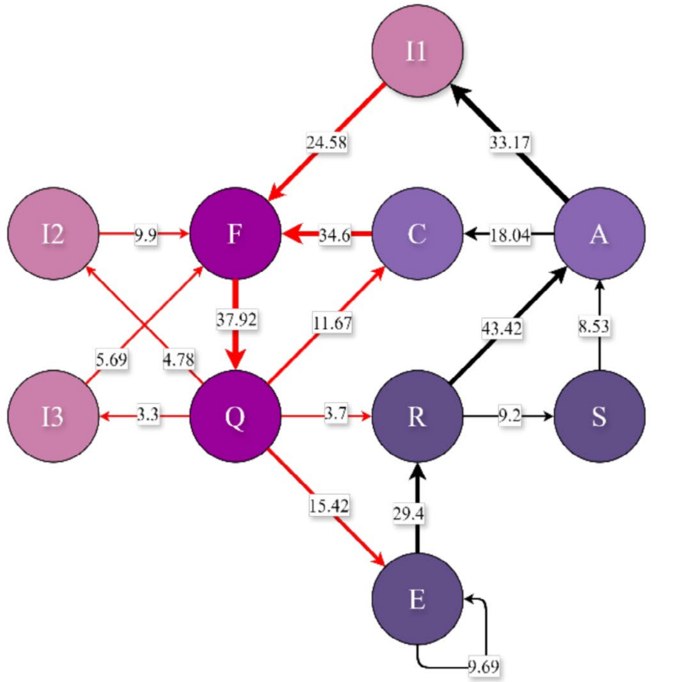
5.5 | Interaction Frequency and Success Rate

The analysis of interaction frequencies (Q_count) within the LLM-ARG experimental group revealed that students interacted with the virtual agent a mean of 20.21 times (SD = 2.50), with a median of 20 interactions (see Table 7). As shown in Figure 9, the distribution was concentrated around this central point; the two most frequent interaction counts were 20 (10 students) and 19 (9 students), together accounting for 68% of the participants (19 out of 28). The full range of interactions spanned from a minimum of 14 to a maximum of 30, indicating that whilst most students exhibited moderate engagement, a small subgroup of learners interacted more extensively with the agent.

5.6 | Problem-Solving Success

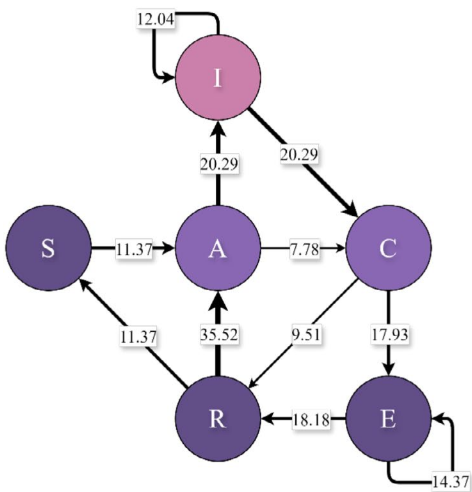
In this study, problem-solving success was quantified using a success rate (Q_success_rate), calculated as the proportion of tasks a student answered correctly after interacting with the LLM-based agent. This approach, adapted from Liang et al. (2024), operationalizes success based on the outcomes of these supported interactions. Overall, the system demonstrated robust efficacy, with students achieving an average success rate of 0.88 (SD = 0.12). The data showed a strong tendency towards successful outcomes: the median success rate was 0.83, and 75% of students achieved a success rate of 0.83 or higher. Notably, at least 25% of the students achieved a perfect 1.00 success rate, as their failure rate was zero.

These results indicate that the LLM-ARG system effectively supported most students in reaching correct solutions. However, the variability in outcomes, with the minimum success rate recorded at 0.62, suggests that some students still faced challenges, particularly with the more complex tasks presented in the later stages



Code	Code name
E	Exploring
R	Reading learning materials
S	Skipping learning materials
A	Answering question
C	Answering the question correctly
I1	Answering the question incorrectly for the first time
I2	Answering the question incorrectly for the second time
I3	Answering the question incorrectly for the third time
F	Choosing to receive feedback from V
Q	Q&A with V

FIGURE 7 | Behavioural transition diagram of the experimental group.



Code	Code name
E	Exploring
R	Reading learning materials
S	Skipping learning materials
A	Answering question
C	Answering the question correctly
I	Answering the question incorrectly

FIGURE 8 | Behavioural transition diagram of the control group.

of the game. The fact that a substantial proportion of students (25%) were able to use the agent to achieve a perfect success rate underscores the system's potential for scaffolding learners towards mastery of foundational content.

5.7 | Interaction With Virtual Agent

To further explore interaction differences, students were stratified into two groups based on a median split of the interaction frequency (Median $=20$): a low-interaction group (Q_count ≤ 20, $n=20$) and a high-interaction group (Q_count >20, $n=8$). As shown in Table 8, an inverse relationship between interaction frequency and immediate success was observed: the low-interaction group achieved higher success ($M=0.92$, SD $=0.09$) than the high-interaction group ($M=0.79$, SD $=0.14$), $t=-2.375$, $p=0.041$.

The lag sequential analysis indicated that high-interaction students engaged in iterative feedback loops, reflecting exploratory learning, whilst low-interaction students employed more goal-directed strategies, achieving greater immediate success. These patterns suggest that interaction frequency alone does not determine learning outcomes, highlighting the importance of different learning strategies.

5.8 | Correlation Analysis

To gain deeper insights into the relationships amongst interaction frequency, metacognitive skills, and academic performance, we conducted a Pearson correlation analysis. Table 9 presents the Pearson correlation matrix amongst key variables, including interaction frequency (Q_count), success rate, metacognitive pretest and posttest scores, as well as academic performance measures (pretest and posttest scores). The results revealed several significant associations between key variables. Notably, pretest and posttest scores were highly positively correlated ($r=0.646$, $p<0.001$), indicating that better initial academic performance led to sustained high performance. Similarly, metacognitive pretest scores were significantly correlated with both pretest scores ($r=0.702$, $p<0.001$) and posttest scores ($r=0.637$, $p<0.001$), suggesting that metacognitive awareness positively influences academic success.

TABLE 7 | Descriptive statistics of key metrics.

Metric	Mean	SD	Min	Max	25%	50%
Q_count	20.21	2.50	14	30	19	20
Success rate	88	12	62	100	83	83
Failure rate	12	12	0	38	0	17

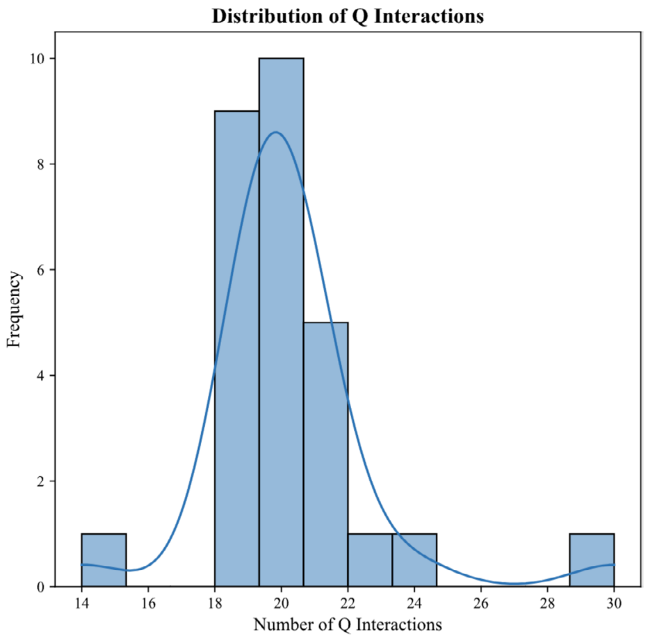


FIGURE 9 | Distribution of Q interactions.

Furthermore, a strong positive correlation was observed between metacognitive pretest and posttest scores ($r=0.723$, $p<0.001$), reflecting the continuity of metacognitive competence throughout the intervention. These findings align with prior research emphasising the critical role of metacognitive regulation in promoting academic achievement.

However, it is noteworthy that interaction frequency (Q_count) and success rate did not exhibit a significant correlation ($r=-0.169$, $p=0.389$), indicating that frequent interactions do not necessarily translate to higher success. Interestingly, a marginally significant negative correlation was observed between interaction frequency and metacognitive pretest scores ($r=-0.367$, $p=0.055$), suggesting that students with higher metacognitive skills tend to interact less frequently, likely due to more efficient problem-solving strategies.

6 | Discussion

In this study, we developed an ARG system based on a personalised LLM virtual agent, known as LLM-ARG. This system supports student interaction with the virtual agent within a gaming environment, providing three types of adaptive feedback (Answer, Task-Oriented, and Achievement Feedback) to foster autonomous learning and interdisciplinary inquiry capabilities. The study examined the effectiveness of the LLM-ARG method compared to the C-ARG method in interdisciplinary STEM educational activities. The results indicate that the LLM-ARG method enhances students' learning achievements and

TABLE 8 | Group comparison of high vs. low interaction based on a median split.

Group	*N*	Q_count (mean)	Success rate (mean)	Standard deviation (success rate)	*t*-value	*p*
High interaction	8	22.63	0.79	0.14	−2.375	0.041
Low interaction	20	19.25	0.92	0.09		

TABLE 9 | Pearson correlation matrix amongst key variables.

	Q_count	Success_rate	Pretest	Posttest	Meta_pre	Meta_post
Q_count	1					
Success_rate	−0.169	1				
Pretest	−0.181	0.314	1			
Posttest	−0.320	0.357	0.646***	1		
Meta_pre	−0.367	0.087	0.702***	0.637***	1	
Meta_post	−0.109	0.096	0.490**	0.305	0.723***	1

**$p<0.01$.
***$p<0.001$.

metacognitive awareness whilst also promoting positive behavioural engagement.

In terms of learning achievements, this study found that students in the experimental group statistically significantly outperformed those in the control group, indicating that the LLM-powered virtual agent contributed to improved learning outcomes. This result can be interpreted through the lens of Cognitive Load Theory (Sweller 1994). The complex, interdisciplinary nature of the STEM tasks could impose a high extraneous cognitive load on students. By providing real-time, explanatory feedback and answering student-initiated questions, the LLM-ARG likely mitigated this load, allowing students to allocate more cognitive resources to problem-solving and knowledge construction. This finding aligns with a growing body of research that positions LLMs not just as information providers, but as tools for scaffolding higher-order thinking skills (Lee et al. 2024). Whilst other studies have shown LLMs supporting critical judgement (Aprin et al. 2024) or ethical problem-solving (Hu 2024), our study extends this understanding to the context of interdisciplinary STEM education. It suggests that an LLM's capacity for contextualised feedback is particularly effective in helping students apply and integrate complex knowledge, which is consistent with other work showing beneficial impacts when integrating gaming with LLM-supported learning (Goslen et al. 2024; Xu et al. 2024). Overall, the findings suggest that integrating LLM virtual agents into ARGs can provide more meaningful learning support for students' academic development in STEM.

In terms of metacognitive awareness, a significant enhancement in the experimental group was observed. This finding strongly supports the role of the LLM-ARG as a tool for scaffolding SRL (Zimmerman 1995). The results are consistent with previous research demonstrating that LLM-supported learning tools can improve students' abilities to understand and manage their cognitive states during the learning process (Goslen et al. 2024; Meyer et al. 2024). The personalised feedback and guidance from the virtual agent likely encouraged students to engage in key SRL processes such as planning, monitoring, and self-evaluation, which are foundational to metacognition (Lim et al. 2023; Maghsudi et al. 2021). Through real-time feedback and adaptive support, the LLM-ARG assists students in reflecting and adjusting throughout the learning process. This mechanism encourages students to become more aware of their learning states, thereby enhancing their self-management and metacognitive skills in future learning tasks. This result highlights the potential of LLM-ARG as a valuable tool in educational contexts that demand a high degree of student self-regulation, such as the interdisciplinary STEM environment explored in this study.

In terms of learning behaviours, lag sequential analysis revealed that the experimental group engaged in productive, iterative learning cycles. Significant behavioural paths showed students interacting with the LLM virtual agent after both correct and incorrect answers ($C \rightarrow F$, $I1 \rightarrow F$), which led to subsequent Q&A dialogues ($F \rightarrow Q$) that resulted in higher accuracy ($Q \rightarrow C$). This contrasts sharply with the control group, whose behaviours were more linear and lacked these rich feedback loops. This suggests that the LLM virtual agent acted as a key scaffold for SRL. By providing timely, explanatory feedback, the agent helped students monitor their understanding and control their learning strategies immediately following an error, prompting them to re-engage more actively with the learning tasks. Although the LLM-ARG encouraged these positive behaviours, students in the experimental group, like those in the control group, still showed a tendency to skip some materials, reflecting a persistent challenge in self-directed learning. Nevertheless, the finding that the agent's instant feedback and continuous interaction promoted more systematic engagement aligns with research on the role of feedback in promoting positive learning behaviours (Duggal et al. 2021; Huang et al. 2023), further validating the potential of LLM-driven systems to support active learning processes.

Analysis of interaction patterns in the LLM-ARG group showed that most students demonstrated moderate engagement, whilst a subgroup engaged intensively through iterative query–revision loops. This aligns with metacognitive refinement strategies in self-regulated learning (Qiao et al. 2022) and was accompanied by an inverse relationship between interaction frequency and immediate success. Low-frequency interactors adopted a more performance-oriented strategy, efficiently resolving problems for higher immediate success, whereas high-frequency interactors followed a mastery-oriented approach, using iterative feedback loops to refine understanding. These findings highlight the system's adaptability to diverse learning strategies and suggest that the value of an AI tutor should be assessed not only by efficiency but also by its ability to support both exploratory and goal-directed learning.

The correlation analysis provides further insight into the relationship between learning behaviours and outcomes. As noted, interaction frequency did not significantly correlate with immediate problem-solving success. In contrast, metacognitive competence emerged as a much more reliable predictor of academic performance. Specifically, pre-intervention metacognitive awareness showed strong, significant positive correlations with both pretest ($r = 0.702$, $p < 0.001$) and posttest academic scores ($r = 0.637$, $p < 0.001$). This aligns with existing research highlighting that students with stronger metacognitive skills tend to perform better academically (Haataja et al. 2022; Katsantonis 2025). This reinforces the interpretation from our behavioural analysis: it is not the quantity of interactions that predicts success, but the underlying quality of a student's learning regulation. The marginally significant negative correlation between interaction frequency and pre-intervention metacognitive awareness ($r = -0.367$, $p = 0.055$) further supports this, suggesting that students who already possess stronger metacognitive skills may adopt more efficient, strategic learning behaviours that require fewer interactions (Liang et al. 2024). Ultimately, these findings underscore the importance of fostering metacognitive competence as a key driver of academic success in technology-enhanced learning environments.

Whilst the LLM-ARG system effectively supported learning progression, the variability in student performance was notable. Although the direct correlation between interaction frequency and success rate was not statistically significant ($r = -0.169$, $p = 0.389$), the comparison between the low- and high-interaction groups revealed a significant difference in their strategies and immediate outcomes. This highlights that a simple linear relationship does not capture the complexity of the learning process and that frequent interactions do not necessarily result in better learning outcomes (Li et al. 2022). As discussed, students appeared to adopt two different learning strategies. Students with higher pre-intervention metacognitive skills tended to employ a more efficient, streamlined problem-solving approach, contributing to higher success rates despite fewer interactions. In contrast, the high-interaction students engaged in an exploratory approach. Whilst this exploratory strategy is valuable for building understanding, it may also increase cognitive load in the short term, requiring a careful balance between exploration and efficiency to optimise learning in complex problem-solving environments (Sankaranarayanan et al. 2024; Seufert et al. 2024).

7 | Conclusions

This study demonstrates that the LLM-ARG system effectively supports adaptive game-based STEM education by balancing exploratory refinement with goal-oriented efficiency. Through LLM-powered scaffolding, the system dynamically facilitates both structured learning and open-ended inquiry, fostering a cognitively coherent and efficient learning experience. The positive correlation between metacognitive competence and academic success highlights the critical role of integrating metacognitive support mechanisms within ARG frameworks to enhance learning outcomes. Moreover, the system's ability to balance engagement and efficacy by accommodating diverse learning strategies further reinforces its educational value.

Despite these achievements, several limitations remain. The study was limited to interdisciplinary STEM education with a relatively small sample size and short duration, potentially affecting the generalizability of findings. Future research should expand the sample size and explore the system's applicability across diverse disciplines and student populations. Additionally, whilst we theorised that the LLM agent helps manage cognitive load, this was not directly measured. Future work could explicitly investigate how the system impacts students' cognitive load during complex problem-solving tasks. This study focused on interaction frequency rather than qualitative strategies. The finding that high-frequency interaction did not yield higher immediate success suggests that the quality of inquiries is critical. High-frequency interactions likely reflect exploratory learning, but their self-regulation efficiency remains unclear. Future research should analyse dialogue content more closely to examine how different strategies affect learning outcomes. The observation that students used both efficient and exploratory strategies opens a new avenue for investigation, particularly regarding their long-term effects on conceptual understanding versus problem-solving speed, and how LLM-driven systems can support flexible strategy switching based on learning goals. By further exploring these nuances, future research can help unlock the full potential of LLMs to support diverse learners in dynamic educational contexts.

Author Contributions

Minkai Wang: conceptualization, methodology, software, data curation, formal analysis, writing – original draft, writing – review and editing. **Jingdong Zhu:** conceptualization, methodology. **Gwo-Jen Hwang:** conceptualization, writing – review and editing. **Shao-Chen Chang:** methodology, writing – review and editing. **Qi-Fan Yang:** methodology, writing – review and editing. **Di Zhang:** conceptualization, methodology, funding acquisition, writing – original draft, writing – review and editing.

Acknowledgements

This work was supported by the National Social Science Foundation of China, Youth Project in Education. The project is titled 'A Study on STEM Curriculum Design for Artificial Intelligence Literacy Enhancement of Elementary and Middle School Students in the New Era' [Project No. CCA230341].

Ethics Statement

This study was conducted under ethical guidelines. Each participant signed an informed consent form, and to ensure confidentiality, students' personal identifiers were removed prior to data processing.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Al-Abri, A. 2025. "Exploring ChatGPT as a Virtual Tutor: A Multi-Dimensional Analysis of Large Language Models in Academic Support." *Education and Information Technologies* 30, no. 12: 1–36. https://doi.org/10.1007/s10639-025-13484-x.

Aprin, F., P. Peters, and H. U. Hoppe. 2024. "The Effectiveness of a Virtual Learning Companion for Supporting the Critical Judgment of Social Media Content." *Education and Information Technologies* 29, no. 10: 12797–12830. https://doi.org/10.1007/s10639-023-12275-6.

Bainbridge, K., V. Shute, S. Rahimi, et al. 2022. "Does Embedding Learning Supports Enhance Transfer During Game-Based Learning?" *Learning and Instruction* 77: 101547. https://doi.org/10.1016/j.learninstruc.2021.101547.

Bibauw, S., T. François, and P. Desmet. 2019. "Discussing With a Computer to Practice a Foreign Language: Research Synthesis and Conceptual Framework of Dialogue-Based CALL." *Computer Assisted Language Learning* 32, no. 8: 827–877. https://doi.org/10.1080/09588221.2018.1535508.

Chang, Y., X. Wang, J. Wang, et al. 2024. "A Survey on Evaluation of Large Language Models." *ACM Transactions on Intelligent Systems and Technology* 15, no. 3: 1–45. https://doi.org/10.1145/3641289.

Chen, H., H. W. Park, and C. Breazeal. 2020. "Teaching and Learning With Children: Impact of Reciprocal Peer Learning With a Social Robot on Children's Learning and Emotive Engagement." *Computers & Education* 150: 103836. https://doi.org/10.1016/j.compedu.2020.103836.

Chen, X., Z. Hu, and C. Wang. 2024. "Empowering Education Development Through AIGC: A Systematic Literature Review." *Education and Information Technologies* 29: 17485–17537. https://doi.org/10.1007/s10639-024-12549-7.

Chen, Y., S. Jensen, L. J. Albert, S. Gupta, and T. Lee. 2023. "Artificial Intelligence (AI) Student Assistants in the Classroom: Designing

Chatbots to Support Student Success." *Information Systems Frontiers* 25, no. 1: 161–182. https://doi.org/10.1007/s10796-022-10291-4.

Chernikova, O., N. Heitzmann, M. Stadler, D. Holzberger, T. Seidel, and F. Fischer. 2020. "Simulation-Based Learning in Higher Education: A Meta-Analysis." *Review of Educational Research* 90, no. 4: 499–541. https://doi.org/10.3102/0034654320933544.

Chou, C.-M., T.-C. Shen, T.-C. Shen, and C.-H. Shen. 2022. "Influencing Factors on Students' Learning Effectiveness of AI-Based Technology Application: Mediation Variable of the Human-Computer Interaction Experience." *Education and Information Technologies* 27, no. 6: 8723–8750. https://doi.org/10.1007/s10639-021-10866-9.

Christopoulos, A., S. Mystakidis, E. Cachafeiro, and M.-J. Laakso. 2023. "Escaping the Cell: Virtual Reality Escape Rooms in Biology Education." *Behaviour & Information Technology* 42, no. 9: 1434–1451. https://doi.org/10.1080/0144929X.2022.2079560.

De Beer, K., and T. Bothma. 2016. "Alternate Reality Games (ARG) as Innovative Digital Information Sources." *Library Hi Tech* 34, no. 3: 433–453. https://doi.org/10.1108/LHT-02-2016-0018.

Dokukina, I., and J. Gumanova. 2020. "The Rise of Chatbots—New Personal Assistants in Foreign Language Learning." *Procedia Computer Science* 169: 542–546. https://doi.org/10.1016/j.procs.2020.02.212.

Du, Q. 2025. "How Artificially Intelligent Conversational Agents Influence EFL Learners' Self-Regulated Learning and Retention." *Education and Information Technologies*: 1–67. https://doi.org/10.1007/s10639-025-13602-9.

Duggal, K., L. R. Gupta, and P. Singh. 2021. "Gamification and Machine Learning Inspired Approach for Classroom Engagement and Learning." *Mathematical Problems in Engineering* 2021: 1–18. https://doi.org/10.1155/2021/9922775.

El Hajji, M., T. Ait Baha, A. Berka, H. Ait Nacer, H. El Aouifi, and Y. Es-Saady. 2025. "An Architecture for Intelligent Tutoring in Virtual Reality: Integrating LLMs and Multimodal Interaction for Immersive Learning." *Information (Basel)* 16, no. 7: 556. https://doi.org/10.3390/info16070556.

Elsom, S., M. Westacott, C. Stieler-Hunt, S. Glencross, and K. Rutter. 2023. "Finding Resources, Finding Friends: Using an Alternate Reality Game for Orientation and Socialisation in a University Enabling Program." *Interactive Learning Environments* 31, no. 5: 2635–2649. https://doi.org/10.1080/10494820.2021.1894181.

English, L. D. 2023. "Ways of Thinking in STEM-Based Problem Solving." *ZDM Mathematics Education* 55, no. 7: 1219–1230. https://doi.org/10.1007/s11858-023-01474-7.

Essel, H. B., D. Vlachopoulos, A. Tachie-Menson, E. E. Johnson, and P. K. Baah. 2022. "The Impact of a Virtual Teaching Assistant (Chatbot) on Students' Learning in Ghanaian Higher Education." *International Journal of Educational Technology in Higher Education* 19, no. 1: 57. https://doi.org/10.1186/s41239-022-00362-6.

Ferguson, C. J. 2009. "An Effect Size Primer: A Guide for Clinicians and Researchers." *Professional Psychology: Research and Practice* 40, no. 5: 532–538. https://doi.org/10.1037/a0015808.

Fountoukidou, S., J. Ham, U. Matzat, and C. Midden. 2019. "Effects of an Artificial Agent as a Behavioral Model on Motivational and Learning Outcomes." *Computers in Human Behavior* 97: 84–93. https://doi.org/10.1016/j.chb.2019.03.013.

Gao, X., P. Li, J. Shen, and H. Sun. 2020. "Reviewing Assessment of Student Learning in Interdisciplinary STEM Education." *International Journal of STEM Education* 7, no. 1: 24. https://doi.org/10.1186/s40594-020-00225-4.

Gilliam, M., P. Jagoda, C. Fabiyi, et al. 2017. "Alternate Reality Games as an Informal Learning Tool for Generating STEM Engagement Among Underrepresented Youth: A Qualitative Evaluation of the Source." *Journal of Science Education and Technology* 26, no. 3: 295–308. https://doi.org/10.1007/s10956-016-9679-4.

Glazewski, K. D., and P. A. Ertmer. 2020. "Fostering Complex Problem Solving for Diverse Learners: Engaging an Ethos of Intentionality Toward Equitable Access." *Educational Technology Research and Development* 68, no. 2: 679–702. https://doi.org/10.1007/s11423-020-09762-9.

Goslen, A., Y. J. Kim, J. Rowe, and J. Lester. 2024. "LLM-Based Student Plan Generation for Adaptive Scaffolding in Game-Based Learning Environments." *International Journal of Artificial Intelligence in Education* 35: 533–558. https://doi.org/10.1007/s40593-024-00421-1.

Guo, X.-R., S.-Y. Liu, S.-Y. Gong, Y. Cao, J. Wang, and Y. Fang. 2024. "Promoting Math Learning in Educational Games With Virtual Companions Providing Learning Supports." *Education and Information Technologies* 29: 22341–22370. https://doi.org/10.1007/s10639-024-12741-9.

Haataja, E., M. Dindar, J. Malmberg, and S. Järvelä. 2022. "Individuals in a Group: Metacognitive and Regulatory Predictors of Learning Achievement in Collaborative Learning." *Learning and Individual Differences* 96: 102146. https://doi.org/10.1016/j.lindif.2022.102146.

Habibi, R., S. S. Maram, J. Pfau, et al. 2022. "A Data-Driven Design of AR Alternate Reality Games to Measure Resilience." In *HCI in Games*, edited by X. Fang, vol. 13334, 586–604. Springer International Publishing.

Han, I., H. Ji, S. Jin, and K. Choi. 2025. "Mobile-Based Artificial Intelligence Chatbot for Self-Regulated Learning in a Hybrid Flipped Classroom." *Journal of Computing in Higher Education*: 1–25. https://doi.org/10.1007/s12528-025-09434-8.

Hew, K. F., W. Huang, J. Du, and C. Jia. 2023. "Using Chatbots to Support Student Goal Setting and Social Presence in Fully Online Activities: Learner Engagement and Perceptions." *Journal of Computing in Higher Education* 35, no. 1: 40–68. https://doi.org/10.1007/s12528-022-09338-x.

Hou, H.-T., Y.-S. Fang, and J. T. Tang. 2023. "Designing an Alternate Reality Board Game With Augmented Reality and Multi-Dimensional Scaffolding for Promoting Spatial and Logical Ability." *Interactive Learning Environments* 31, no. 7: 4346–4366. https://doi.org/10.1080/10494820.2021.1961810.

Hu, B., L. Zheng, J. Zhu, L. Ding, Y. Wang, and X. Gu. 2024. "Teaching Plan Generation and Evaluation With GPT-4: Unleashing the Potential of LLM in Instructional Design." *IEEE Transactions on Learning Technologies* 17: 1471–1485. https://doi.org/10.1109/TLT.2024.3384765.

Hu, Y.-H. 2024. "Improving Ethical Dilemma Learning: Featuring Thinking Aloud Pair Problem Solving (TAPPS) and AI-Assisted Virtual Learning Companion." *Education and Information Technologies* 29: 22969–22990. https://doi.org/10.1007/s10639-024-12754-4.

Huang, A. Y. Q., O. H. T. Lu, and S. J. H. Yang. 2023. "Effects of Artificial Intelligence-Enabled Personalized Recommendations on Learners' Learning Engagement, Motivation, and Outcomes in a Flipped Classroom." *Computers & Education* 194: 104684. https://doi.org/10.1016/j.compedu.2022.104684.

Jagoda, P., M. Gilliam, P. McDonald, and C. Russell. 2015. "Worlding Through Play: Alternate Reality Games, Large-Scale Learning, and the Source." *American Journal of Play* 8, no. 1: 74–100.

Jesionkowska, J., F. Wild, and Y. Deval. 2020. "Active Learning Augmented Reality for STEAM Education—A Case Study." *Education in Science* 10, no. 8: 198. https://doi.org/10.3390/educsci10080198.

Jin, S. H., K. Im, M. Yoo, I. Roll, and K. Seo. 2023. "Supporting Students' Self-Regulated Learning in Online Learning Using Artificial Intelligence Applications." *International Journal of Educational Technology in Higher Education* 20, no. 1: 37. https://doi.org/10.1186/s41239-023-00406-5.

Kasneci, E., K. Sessler, S. Küchemann, et al. 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.

Kassutto, S. M., C. Baston, and C. Clancy. 2021. "Virtual, Augmented, and Alternate Reality in Medical Education: Socially Distanced but Fully Immersed." *ATS Scholar* 2, no. 4: 651–664. https://doi.org/10.34197/ats-scholar.2021-0002RE.

Katsantonis, I. G. 2025. "Typologies of Teaching Strategies in Classrooms and Students' Metacognition and Motivation: a Latent Profile Analysis of the Greek PISA 2018 Data." *Metacognition and Learning* 20, no. 1: 1–22. https://doi.org/10.1007/s11409-024-09410-0.

Krassmann, A. L., F. B. Nunes, M. Bessa, L. M. R. Tarouco, and M. Bercht. 2019. "Virtual Companions and 3D Virtual Worlds: Investigating the Sense of Presence in Distance Education." In *Learning and Collaboration Technologies. Ubiquitous and Virtual Environments for Learning and Collaboration*, edited by P. Zaphiris and A. Ioannou, 175–192. Springer International Publishing.

Kuhail, M. A., N. Alturki, S. Alramlawi, and K. Alhejori. 2023. "Interacting With Educational Chatbots: A Systematic Review." *Education and Information Technologies* 28, no. 1: 973–1018. https://doi.org/10.1007/s10639-022-11177-3.

Lai, C.-L., and G.-J. Hwang. 2015. "Effects of Mobile Learning Time on Students' Conception of Collaboration, Communication, Complex Problem–Solving, Meta–Cognitive Awareness and Creativity." *International Journal of Mobile Learning and Organisation* 8: 276–291. https://doi.org/10.1504/IJMLO.2014.067029.

Lee, H. Y., P. H. Chen, W. S. Wang, Y. M. Huang, and T. T. Wu. 2024. "Empowering ChatGPT With Guidance Mechanism in Blended Learning: Effect of Self-Regulated Learning, Higher-Order Thinking Skills, and Knowledge Construction." *International Journal of Educational Technology in Higher Education* 21, no. 1: 16. https://doi.org/10.1186/s41239-024-00447-4.

Lee, Y.-F., G.-J. Hwang, and P.-Y. Chen. 2022. "Impacts of an AI-Based Chabot on College Students' After-Class Review, Academic Performance, Self-Efficacy, Learning Attitude, and Motivation." *Educational Technology Research and Development* 70, no. 5: 1843–1865. https://doi.org/10.1007/s11423-022-10142-8.

Li, X., X. Lin, F. Zhang, and Y. Tian. 2022. "What Matters in Online Education: Exploring the Impacts of Instructional Interactions on Learning Outcomes." *Frontiers in Psychology* 12: 792464. https://doi.org/10.3389/fpsyg.2021.792464.

Li, Y., K. Wang, Y. Xiao, and J. E. Froyd. 2020. "Research and Trends in STEM Education: A Systematic Review of Journal Publications." *International Journal of STEM Education* 7, no. 1: 11. https://doi.org/10.1186/s40594-020-00207-6.

Liang, H., T. Hsu, and G. Hwang. 2021. "Promoting Children's Inquiry Performances in Alternate Reality Games: A Mobile Concept Mapping-Based Questioning Approach." *British Journal of Educational Technology* 52, no. 5: 2000–2019. https://doi.org/10.1111/bjet.13095.

Liang, H., G. Hwang, T. Hsu, and J. Yeh. 2024. "Effect of an AI-Based Chatbot on Students' Learning Performance in Alternate Reality Game-Based Museum Learning." *British Journal of Educational Technology* 55: 2315–2338. https://doi.org/10.1111/bjet.13448.

Liao, C.-W., C.-H. Chen, and S.-J. Shih. 2019. "The Interactivity of Video and Collaboration for Learning Achievement, Intrinsic Motivation, Cognitive Load, and Behavior Patterns in a Digital Game-Based Learning Environment." *Computers & Education* 133: 43–55. https://doi.org/10.1016/j.compedu.2019.01.013.

Lim, L., M. Bannert, J. Van Der Graaf, et al. 2023. "Effects of Real-Time Analytics-Based Personalized Scaffolds on Students' Self-Regulated Learning." *Computers in Human Behavior* 139: 107547. https://doi.org/10.1016/j.chb.2022.107547.

Lin, C.-J., and H. Mubarok. 2021. "Learning Analytics for Investigating the Mind Map-Guided AI Chatbot Approach in an EFL Flipped Speaking Classroom." *Educational Technology & Society* 24, no. 4: 16–35.

Lin, L., P. Ginns, T. Wang, and P. Zhang. 2020. "Using a Pedagogical Agent to Deliver Conversational Style Instruction: What Benefits Can You Obtain?" *Computers & Education* 143: 103658. https://doi.org/10.1016/j.compedu.2019.103658.

Lin, Y.-J., and H. Wang. 2021. "Using Virtual Reality to Facilitate Learners' Creative Self-Efficacy and Intrinsic Motivation in an EFL Classroom." *Education and Information Technologies* 26, no. 4: 4487–4505. https://doi.org/10.1007/s10639-021-10472-9.

Maghsudi, S., A. Lan, J. Xu, and M. van der Schaar. 2021. "Personalized Education in the Artificial Intelligence Era: What to Expect Next." *IEEE Signal Processing Magazine* 38, no. 3: 37–50. https://doi.org/10.1109/MSP.2021.3055032.

Meekaew, N., and W. Ketpichainarong. 2021. "The Effects of Augmented Reality-Facilitated Mobile Game-Based Learning on the Diversity of Life for Promoting Learning at the Natural History Museum." *International Journal of Mobile Learning and Organisation* 15, no. 3: 282. https://doi.org/10.1504/IJMLO.2021.116509.

Meyer, J., T. Jansen, R. Schiller, et al. 2024. "Using LLMs to Bring Evidence-Based Feedback Into the Classroom: AI-Generated Feedback Increases Secondary Students' Text Revision, Motivation, and Positive Emotions." *Computers and Education: Artificial Intelligence* 6: 100199. https://doi.org/10.1016/j.caeai.2023.100199.

Moon, J., and J. Ryu. 2021. "The Effects of Social and Cognitive Cues on Learning Comprehension, Eye-Gaze Pattern, and Cognitive Load in Video Instruction." *Journal of Computing in Higher Education* 33, no. 1: 39–63. https://doi.org/10.1007/s12528-020-09255-x.

Oker, A., F. Pecune, and C. Declercq. 2020. "Virtual Tutor and Pupil Interaction: A Study of Empathic Feedback as Extrinsic Motivation for Learning." *Education and Information Technologies* 25, no. 5: 3643–3658. https://doi.org/10.1007/s10639-020-10123-5.

Ooi, K.-B., G. W.-H. Tan, M. Al-Emran, et al. 2023. "The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions." *Journal of Computer Information Systems* 65, no. 1: 1–32. https://doi.org/10.1080/08874417.2023.2261010.

Örnebring, H. 2007. "Alternate Reality Gaming and Convergence Culture: The Case of Alias." *International Journal of Cultural Studies* 10, no. 4: 445–462. https://doi.org/10.1177/1367877907083079.

Pettersson, F. 2021. "Understanding Digitalization and Educational Change in School by Means of Activity Theory and the Levels of Learning Concept." *Education and Information Technologies* 26, no. 1: 187–204. https://doi.org/10.1007/s10639-020-10239-8.

Qian, Y., and I. Choi. 2023. "Tracing the Essence: Ways to Develop Abstraction in Computational Thinking." *Educational Technology Research and Development* 71, no. 3: 1055–1078. https://doi.org/10.1007/s11423-022-10182-0.

Qiao, S., S. K. W. Chu, X. Shen, and S. S. S. Yeung. 2022. "The Impact of an Online Gamified Approach Embedded With Self-Regulated Learning Support on Students' Reading Performance and Intrinsic Motivation: A Randomized Controlled Trial." *Journal of Computer Assisted Learning* 38, no. 5: 1379–1393.

Richey, J. E., J. M. L. Andres-Bray, M. Mogessie, et al. 2019. "More Confusion and Frustration, Better Learning: The Impact of Erroneous Examples." *Computers & Education* 139: 173–190. https://doi.org/10.1016/j.compedu.2019.05.012.

Sankaranarayanan, R., M. Yang, and K. Kwon. 2024. "Exploring the Role of a Microlearning Instructional Approach in an Introductory Database Programming Course: an Exploratory Case Study." *Journal of Computing in Higher Education* 37, no. 3: 1–864. https://doi.org/10.1007/s12528-024-09408-2.

Schlimbach, R., B. Khosrawi-Rad, T. Lange, T. Strohmann, and S. Robra-Bissantz. 2024. "Design Knowledge for Virtual Learning Companions From a Value-Centered Perspective." *Communications of the Association for Information Systems* 54, no. 1: 293–330. https://doi.org/10.17705/1CAIS.05411.

Schraw, G., and R. S. Dennison. 1994. "Assessing Metacognitive Awareness." *Contemporary Educational Psychology* 19, no. 4: 460–475. https://doi.org/10.1006/ceps.1994.1033.

Seo, K., J. Tang, I. Roll, S. Fels, and D. Yoon. 2021. "The Impact of Artificial Intelligence on Learner–Instructor Interaction in Online Learning." *International Journal of Educational Technology in Higher Education* 18, no. 1: 54. https://doi.org/10.1186/s41239-021-00292-9.

Seufert, T. 2018. "The Interplay Between Self-Regulation in Learning and Cognitive Load." *Educational Research Review* 24: 116–129. https://doi.org/10.1016/j.edurev.2018.03.004.

Seufert, T., V. Hamm, A. Vogt, and V. Riemer. 2024. "The Interplay of Cognitive Load, Learners' Resources and Self-Regulation." *Educational Psychology Review* 36, no. 2: 50. https://doi.org/10.1007/s10648-024-09890-1.

Sewell, J. L., J. L. Bowen, O. T. Cate, P. S. O'Sullivan, B. Shah, and C. K. Boscardin. 2020. "Learning Challenges, Teaching Strategies, and Cognitive Load: Insights From the Experience of Seasoned Endoscopy Teachers." *Academic Medicine* 95, no. 5: 794–802. https://doi.org/10.1097/ACM.0000000000002946.

Sikström, P., C. Valentini, A. Sivunen, and T. Kärkkäinen. 2022. "How Pedagogical Agents Communicate With Students: A Two-Phase Systematic Review." *Computers & Education* 188: 104564. https://doi.org/10.1016/j.compedu.2022.104564.

Sinatra, A. M., K. A. Pollard, B. T. Files, A. H. Oiknine, M. Ericson, and P. Khooshabeh. 2021. "Social Fidelity in Virtual Agents: Impacts on Presence and Learning." *Computers in Human Behavior* 114: 106562. https://doi.org/10.1016/j.chb.2020.106562.

Sofianidis, A., C. Skraparlis, and N. Stylianidou. 2024. "Combining Inquiry, Universal Design for Learning, Alternate Reality Games and Augmented Reality Technologies in Science Education: The IB-ARGI Approach and the Case of Magnetman." *Journal of Science Education and Technology* 33: 928–953. https://doi.org/10.1007/s10956-024-10135-7.

Song, D., M. Rice, and E. Y. Oh. 2019. "Participation in Online Courses and Interaction With a Virtual Agent." *International Review of Research in Open and Distributed Learning* 20, no. 1: 43–62. https://doi.org/10.19173/irrodl.v20i1.3998.

Srisawasdi, N., and P. Panjaburee. 2019. "Implementation of Game-Transformed Inquiry-Based Learning to Promote the Understanding of and Motivation to Learn Chemistry." *Journal of Science Education and Technology* 28, no. 2: 152–164. https://doi.org/10.1007/s10956-018-9754-0.

Stylianidou, N., A. Sofianidis, E. Manoli, and M. Meletiou-Mavrotheris. 2020. ""Helping Nemo!"—Using Augmented Reality and Alternate Reality Games in the Context of Universal Design for Learning." *Education in Science* 10, no. 4: 4. https://doi.org/10.3390/educsci10040095.

Sweller, J. 1994. "Cognitive Load Theory, Learning Difficulty, and Instructional Design." *Learning and Instruction* 4, no. 4: 295–312. https://doi.org/10.1016/0959-4752(94)90003-5.

Tan, A.-L., Y. S. Ong, Y. S. Ng, and J. H. J. Tan. 2023. "STEM Problem Solving: Inquiry, Concepts, and Reasoning." *Science & Education* 32, no. 2: 381–397. https://doi.org/10.1007/s11191-021-00310-2.

Thirunavukarasu, A. J., D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting. 2023. "Large Language Models in Medicine." *Nature Medicine* 29, no. 8: 1930–1940. https://doi.org/10.1038/s41591-023-02448-8.

Tulloch, R., H. Wolfenden, and H. Sercombe. 2021. "Designing Alternate Reality Games for Effective Learning: A Methodology for Implementing Multimodal Persistent Gaming in University Education." *Media Practice and Education* 22, no. 2: 136–152. https://doi.org/10.1080/25741136.2020.1864179.

Vayanou, M., Y. Ioannidis, G. Loumos, and A. Kargas. 2019. "How to Play Storytelling Games With Masterpieces: From Art Galleries to Hybrid Board Games." *Journal of Computers in Education* 6, no. 1: 79–116. https://doi.org/10.1007/s40692-018-0124-y.

Wang, X., H. Pang, M. P. Wallace, Q. Wang, and W. Chen. 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, no. 4: 814–840. https://doi.org/10.1080/09588221.2022.2056203.

Whitton, N., R. Jones, S. Wilson, and P. Whitton. 2014. "Alternate Reality Games as Learning Environments for Student Induction." *Interactive Learning Environments* 22, no. 3: 243–252. https://doi.org/10.1080/10494820.2011.641683.

Xia, Q., T. K. F. Chiu, C. S. Chai, and K. Xie. 2023. "The Mediating Effects of Needs Satisfaction on the Relationships Between Prior Knowledge and Self-Regulated Learning Through Artificial Intelligence Chatbot." *British Journal of Educational Technology* 54, no. 4: 967–986. https://doi.org/10.1111/bjet.13305.

Xu, W., and F. Ouyang. 2022. "A Systematic Review of AI Role in the Educational System Based on a Proposed Conceptual Framework." *Education and Information Technologies* 27, no. 3: 4195–4223. https://doi.org/10.1007/s10639-021-10774-y.

Xu, Y., J. Zhu, M. Wang, F. Qian, Y. Yang, and J. Zhang. 2024. "The Impact of a Digital Game-Based AI Chatbot on Students' Academic Performance, Higher-Order Thinking, and Behavioral Patterns in an Information Technology Curriculum." *Applied Sciences* 14, no. 15: 6418. https://doi.org/10.3390/app14156418.

Yan, L., S. Greiff, Z. Teuber, and D. Gašević. 2024. "Promises and Challenges of Generative Artificial Intelligence for Human Learning." *Nature Human Behaviour* 8, no. 10: 1839–1850. https://doi.org/10.1038/s41562-024-02004-5.

Yang, X., Z. Wang, Q. Wang, K. Wei, K. Zhang, and J. Shi. 2024. "Large Language Models for Automated Q&A Involving Legal Documents: A Survey on Algorithms, Frameworks and Applications." *International Journal of Web Information Systems* 20, no. 4: 413–435. https://doi.org/10.1108/IJWIS-12-2023-0256.

Zhang, L., R. A. Carter Jr., J. D. Basham, and S. Yang. 2022. "Integrating Instructional Designs of Personalized Learning Through the Lens of Universal Design for Learning." *Journal of Computer Assisted Learning* 38, no. 6: 1639–1656. https://doi.org/10.1111/jcal.12725.

Zimmerman, B. J. 1995. "Self-Regulation Involves More Than Metacognition: A Social Cognitive Perspective." *Educational Psychologist* 30, no. 4: 217–221. https://doi.org/10.1207/s15326985ep3004_8.