Journal of ExoTechnology and Education
Vol. 2, No. 1, 2026

Online ISSN: 3079-7950

Author
Ms. Celeste Kotze

Affiliation
Interlink @ Al Yamamah University, Khobar

ORCID link
https://orcid.org/0009-0005-8335-0501

Correspondence to
Ms. Celeste Kotze

E-mail
celeste@interlink.edu

Dates
Received: 10 December 2026
Revised: 28 February 2026
Accepted: 1 March 2026
Published: 24 May 2026

How to cite
Kotze, C. (2026) 'Integrating agentic AI and immersive technologies into standardised and adaptive CEFR-Aligned ESL learning environments in higher education: A systematic review of prospects and pitfalls,' *jeXed* 2(1). 22 pages. https://doi.org/10.36615/k5bt2774

Article includes
☒ Peer review
☐ Supplementary material

Data availability
☐ Open data set
☒ All data included
☐ On request from author/s
☐ Not available
☐ Not applicable

Editor/s
Dr. Irene Lubbe

Funding
No funding was received for this article.

Conflict of interest
None.

Society with whom we Collaborate
International Conference on Information Communication Technologies in Education

Website: https://icicte.eu/

Integrating Agentic AI and Immersive Technologies into Standardised and Adaptive CEFR-Aligned ESL Learning Environments in Higher Education: A Systematic Review of Prospects and Pitfalls

Abstract

This systematic literature review examines the integration of agentic artificial intelligence (AI) and immersive technologies in English as a second language (ESL) higher education (HE) focused on alignment with the Common European Framework of Reference for Languages (CEFR). Following the PRISMA 2020 guidelines, 42 peer-reviewed studies published between 2010 and 2025 were included. The review introduced a four-criterion classification framework: Autonomy, goal-directed behaviour, dynamic decision-making, and proactivity to distinguish agentic AI systems from adaptive, rule-based, and intelligent tutoring systems. Applying this framework revealed that none of the reviewed studies satisfied all four criteria simultaneously. Only one of the 42 reviewed studies employed a validated, CEFR-aligned measurement instrument, meaning that proficiency gain claims in the remaining literature are inferred rather than verified. The review identified five thematic areas from the thematic analysis: Adaptive learning and personalisation, immersive technologies in ESL education, standardisation and CEFR integration, pedagogical and didactic considerations, and ethical and social implications. The findings suggest a persistent technology-pedagogy asynchrony, whereby immersive environments have advanced faster than AI systems designed to support adaptive instruction within them. Immersive technologies show promise in reducing student anxiety and enrich-

ing contextualised language practice, but learning gains remain mixed and limited to short-term applications.

Interdisciplinary Implications

This review draws on applied linguistics, educational technology, cognitive science, artificial intelligence research, and assessment theory, as its findings carry distinct implications across all five fields. The four-criterion classification framework, autonomy, goal-directed behaviour, dynamic decision-making, and proactivity offers Artificial Intelligence (AI) and computer science researchers a transferable diagnostic tool for distinguishing genuinely agentic systems from the adaptive and rule-based alternatives that dominate current ESL technology design. The technology-pedagogy asynchrony this review identifies, whereby immersive environments outpace the AI systems built to support instruction within them, demands that cognitive scientists revisit cognitive load theory for multimodal, AI-mediated contexts rather than applying frameworks developed for traditional learning environments. For applied linguists and language assessment specialists, the finding that only one of 42 reviewed studies employed a validated, CEFR-aligned measurement instrument exposes a critical methodological deficit that CEFR framework developers and language testing bodies such as Cambridge Assessment must address directly. Education policymakers and institutional leaders must treat the five thematic areas that this review surfaces (adaptive personalisation, immersive technologies, CEFR integration, pedagogy, and ethics) as an interconnected agenda rather than separate procurement decisions. For data governance specialists, the collection of behavioural, affective, and biometric student data within agentic AI and immersive learning environments create regulatory obligations that existing GDPR provisions do not fully cover, requiring purpose-built frameworks developed in collaboration across all these disciplines.

Keywords

Agentic AI, adaptive learning, immersive technologies, the Common European Framework of Reference for Languages (CEFR), English Second Language (ESL), higher education (HE), systematic review, virtual reality

1. Introduction

Recent scholarship provides convincing evidence for the benefits of integrating agentic AI and immersive technologies such as virtual reality (VR), augmented reality (AR), and intelligent virtual assistants in university-level English Second Language (ESL) education (Wang and Huang, 2025; Yu, 2023). Kostopoulos *et al*. (2025) document a positive trajectory from early intelligent tutoring systems to autonomous learning companions capable of customised, real-time guidance. Chiqui Vera *et al.* (2026) have found significant improvements in linguistic competence, particularly in speaking fluency and writing accuracy in higher education (HE) ESL. The envisioned future is one in which highly personalised, adaptable learning systems allow AI to tailor instructional methods, while immersive simulations provide students with realistic, authentic language practice, all aligned with international standards such as the CEFR (Council of Europe, 2020). A critical gap remains in aligning adaptive learning technologies with standardised frameworks.

1.1 Background and Rationale

This creates a pressing need for a comprehensive review of the relevant literature to discern the claims made about agentic AI's potential and effective implementation. Considering the emergence of advanced VR and AR platforms, for instance, has often developed separately from advances in AI systems that can autonomously make decisions about educational mentoring and provide effective student support. Furthermore, while the CEFR provides a conceptual framework for assessing students' language ability at a broad (macro) level of assessment, it does not provide the detailed metrics required to assess the added value that can be obtained by immersive technologies. This review is guided by the following research questions:

1. *To what extent does the existing empirical literature support the effective integration of agentic AI and immersive technologies in developing standard-ised yet adaptive CEFR-aligned ESL learning environments in HE*?
2. *What key challenges and barriers are identified in the literature regarding the successful implementation of agentic AI and immersive technologies in CEFR-aligned ESL learning environments in HE*?

1.2 Significance and Implications of the Review

This review focuses specifically on agentic AI systems characterised by autonomy, goal-directed behaviour, and dynamic decision-making rather than treating all AI-enhanced learning tools as equivalent. The review considers agentic AI and immersive technologies as a combined pedagogical proposition rather than treating them in isolation, reflecting how they are increasingly co-deployed in practice. It also surfaces a persistent tension between how agentic AI is theorised in the literature and what reviewed studies classify as 'agentic.' Finally, the review applies CEFR alignment as an explicit evaluative criterion, assessing not merely whether technologies improve language outcomes but whether those outcomes are linked to internationally recognised proficiency standards.

2. Conceptual Framework

2.1 Overview of Agentic AI

Agentic AI refers to AI systems capable of making independent decisions, adapting learning experiences, and making purposeful changes to the learning environment without constant direct input from a human educator. Kostopoulos *et al.* (2025) describe agentic AI systems as self-contained, goal-based entities designed to operate autonomously with little human intervention and respond dynamically to shifting contexts. Similarly, Goyal (2025) characterises agentic AI as encompassing intelligent systems capable of autonomous, proactive, and goal-focused behaviour through adaptive interactions with their environment. Factors such as processing speed, memory capacity, prior learning, and patterns of attention can serve as vital inputs that shape how the system interprets its interactions with the student. By monitoring these aspects, the system can provide specific learning pathways that target short-term objectives or long-term proficiency goals.

New technological paradigms show that agentic AI often includes social intelligence components as seen in autonomous robots that interact according to social norms (Lin, Jhang and Wang, 2024, p. 6075). Social constructivism provides evidence for how students learn when they interact collaboratively to create their own understanding of knowledge, whereas connectivism explains how learning occurs through the connection of ideas (Zakaria and Ponniah, 2024). The Cognitive Load Theory demonstrates how adaptive systems can manage cognitive load and enhance the efficiency of learning processes (Sweller, Van Merriënboer and Paas, 2019, p. 261). Together, these theories offer a lens through which a standardised framework, such as the CEFR, coupled with an AI system that continuously adapts to individual students' evolving needs, can significantly support cognitive processing by reducing unnecessary load.

The fundamental aspects of the agentic nature of AI in an educational environment are as follows: Continuous adaptability by utilising multiple modalities of user input to create the learning experience; support for both standardised and personalised assessment measures; early detection of learning gaps; facilitation of socially interactive learning opportunities; and preservation of the human pedagogical element, which is a vital source of motivation and contextual grounding (Goyal, 2025; Kostopoulos *et al*., 2025). These features offer flexibility within technology enhanced learning environments, enabling responsive adjustments to changing educational needs, a particularly useful characteristic for ESL courses navigating diverse student profiles and progression trajectories ensuring alignment with frameworks such as CEFR levels for ESL proficiency (Council of Europe, 2020).

2.1.1 Defining Agentic AI: Key Features and Distinctions

It is important to distinguish agentic AI from related but conceptually different systems as the term is applied loosely and, in many cases, inaccurately across the reviewed literature. Three categories of systems are commonly conflated with it:

- A *rule-based system* operates on a set of predetermined conditioned (if) statements meaning it can only generate responses according to those predetermined conditions. For example, presenting a harder question after each correct answer, regardless of overall performance.
- An *adaptive system* takes this further by enabling the system to modify parameters within its rules. Adaptive systems may, for instance, increase the difficulty of questions as a student's score improves. However, there are limits to the effectiveness of adaptive systems. Research indicates that many users of technology-enhanced learning systems report no noticeable benefit from adaptive features, while a few report positive effects (Du Plooy, Casteleijn and Franzsen, 2024). Adaptive systems cannot set its own goals or reconsider its overall instructional approach, while critical mediating factors include students' emotional intelligence and personal preferences, which can sometimes override the intended impact of adaptive features.
- An *intelligent tutoring system* (ITS) models a student's knowledge state within a defined domain and uses that model to guide instruction though typically within a narrowly specified curriculum.

What distinguishes an agentic AI system is that it exhibits four characteristics that must simultaneously be present:

- *Autonomy*: The system initiates a plan of action without direct instruction from the user.
- *Goal-directed behaviour*: It acts towards a learning objective that it has itself generated, selected, or revised, based on its own analysis.
- *Dynamic decision-making*: In constructing a plan of action for a student, the system draws on a continuously evolving model of that student.
- *Proactivity*: It identifies a potential learning gap before it becomes apparent to either the student or the educator.

In practical terms an agentic system does not simply adjust quiz difficulty. It may conclude that a student needs a different modality altogether, restructure the sequence of a session, or flag a motivational issue to the educator all without being instructed to do so. The distinction is not merely taxonomic. It has direct implications for what outcomes one can attribute to 'agentic' interventions in the research literature.

Table 1: Classification of AI system types

Dimension	Rule-based	Adaptive	Intelligent Tutoring	Agentic AI
Decision-making	Fixed if-then rules	Parameter adjustment	Domain inference	Dynamic goal setting
Student modelling	None	Statistical	Knowledge-state model	Multimodal, continuous
Initiative	Reactive	Reactive	Partly proactive	Proactive
CEFR alignment potential	Low	Medium	Medium	High (theoretical)

Note: During data extraction, each reviewed study was assessed against a four-criterion classification framework developed by this review, drawing on the characterisations of agentic AI offered by Kostopoulos et al. (2025) and Goyal (2025). A system was classified as 'agentic' only if the study provided evidence of all four dimensions. Systems meeting only some criteria were classified as 'adaptive' or 'intelligent tutoring' accordingly. This classification was conducted independently by two reviewers during full text assessment, with disagreements resolved through discussion (see Table 1, Appendix A). This operationalisation revealed that most systems described as 'agentic' in the reviewed literature did not satisfy all four criteria this finding constitutes a key contribution of this review.

3. Methodology

3.1 Systematic Review Framework

The systematic review for this study followed PRISMA 2020 guidelines (Page *et al.*, 2021) to ensure that it is both transparent and replicable. The purpose of this systematic review was to examine research on ESL instruction using both immersive technologies and agentic AI in HE, focusing specifically on whether or how these instructional methods align with the CEFR.

3.2 Eligibility Criteria

The review applied the following criteria to choose from studies to be reviewed:

- *Inclusion criteria*: Studies in peer-reviewed English language published from 2010 to 2025, investigating the application of AI or immersive technologies (AR, VR, and simulations), within the HE ESL environment, related to the CEFR (either explicitly stated or implicitly stated). Theoretical and empirical studies were both included.
- *Exclusion criteria*: Excluded were non-peer-reviewed opinion articles, editorial writings, and those studies that were based outside of HE and not based on ESL, nor did they have an AI/immersive technology as their primary focus. Also excluded were duplicate studies and studies where it was impossible to access the full study.

3.3 Information Sources and Search Strategy

The researcher conducted a comprehensive literature search using electronic databases such as Scopus, Web of Science, ERIC, ScienceDirect, Google Scholar, and Elicit (AI-powered research assistant for systematic reviews) to identify relevant studies on AI, immersive technologies (such as VR and AR simulation), and ESL.

The researcher used search terms that were organised into three core themes: 1) 'Agentic AI,' 'AI,' or 'adaptive learning;' 2) 'Immersive Technology,' 'VR,' 'AR,' or 'simulation;' and 3) 'ESL' and 'CEFR.' The researcher changed the search strategy based on the syntax of each database. In addition, the researcher also used backward citation chaining from key articles found in the database search yielding 43 additional records and citations tracking using Google Scholar. The final database search was completed on 15 October 2025.

3.4 Study Selection Process

Study selection followed a three-phase process aligned with PRISMA 2020 recommendations:

Phase 1: Identification

- Database searches delivered 1,247 records.
- Citation chaining yielded 43 additional records in the literature search.
- Total number of records identified: 1,290.
- 255 duplicate records were removed.
- 1,035 unique records remained for screening.

Phase 2: Screening

- Titles and abstracts of 1,035 unique records were screened against eligibility criteria.
- Screening was conducted by one reviewer with AI-assisted prioritisation through Elicit.
- 847 records were excluded for the following reasons:
    - Irrelevant topic or context (n = 521).
    - Wrong educational level (K-12 or non-formal education) (n = 198).
    - Not peer-reviewed (n = 89).
    - Wrong publication type (editorials, commentaries) (n = 39).
- 188 records were deemed potentially eligible, and full-text retrieval was attempted.

Phase 3: Eligibility Assessment

- 15 full texts could not be retrieved.
- 173 full texts were successfully retrieved and assessed for eligibility.
- Two reviewers independently assessed full texts against inclusion criteria.
- Disagreements were resolved through discussion and, when necessary, consultation with a third reviewer.
- 131 full-text articles were excluded for the following reasons:
    - Insufficient focus on agentic AI or immersive technologies (n = 54).
    - No connection to ESL/EFL context (n = 38).
    - Lack of empirical data or theoretical framework (n = 27).
    - Insufficient methodological quality or clarity (n = 12).
- 42 studies were included in the final qualitative synthesis.

A PRISMA 2020 flow diagram documenting the study selection process is provided in Appendix A (Figure 1).

3.5 Data Extraction

Data extraction used a structured template piloted on five studies. Two reviewers extracted data, combined results, and a third reviewer resolved discrepancies. Extracted data included study details, context, methodology, technology focus, key findings, and quality score.

3.6 Quality Appraisal

To assess the quality of the identified studies, a Critical Appraisal Skills Programme (CASP) Checklist adapted for educational technology research (CASP, 2018) with ten criteria was used. Each study was rated across ten criteria, including clarity of research aims, appropriateness of methodology, rigour of data collection and analysis, clarity of findings, and consideration of ethical issues. The studies were classified into three categories based on their CASP scores:

- *High Quality* (n = 26); 8-10 out of 10: Typically, clearly stated research objectives, appropriately designed methodologies, and well-described, significant findings were found in these studies.
- *Medium Quality* (n = 14); 5-7 out of 10: In some cases, the medium quality studies had minor issues with methodology and/or reporting.
- *Lower Quality* (n = 2); 3-4 out of 10: Significant methodological deficiencies, including inadequate sample size and/or poorly defined outcomes existed in all the lower quality studies.

Studies were not excluded based on quality scores alone, instead quality ratings informed the interpretation and synthesis of the findings. A quality appraisal summary is presented in Appendix A (Figure 2).

3.7 Data Synthesis

A narrative investigation followed the thematic analysis of Thomas and Harden (2008). Five major themes emerged: 1) Adaptive learning and personalisation via agentic AI; 2) immersive technologies in ESL education; 3) standardisation and assessment (CEFR integration); 4) pedagogical and didactic considerations; and 5) ethical and social implications.

3.8 Risk of Bias and Methodological Limitations

The review acknowledged the following potential bias: There is a high chance that many studies were excluded because they did not appear in English-language peer-reviewed literature (grey literature was excluded). Technology has been rapidly evolving, and it is possible that some developments have occurred since this review was completed. Additionally, there was a significant variation among the interventions and methods used in the reviewed literature. Using PRISMA, transparent reporting, and systematic extraction support the review's reliability.

3.9 Ethical Considerations

No ethical approval was required, as only published literature was analysed, and the researcher declares no competing interests.

4. Discussion and Analysis

4.1 Thematic Synthesis: Between Promise and Evidence

The analysis found that there were no reviewed studies that met all four criteria for agentic AI at the same time. Studies referring to system(s) as ‘agentic,’ without meeting the four criteria, were treated as a terminological and methodological gap in the synthesis, and not as a reason for excluding them. The narrative synthesis revealed five core themes, highlighting the tension between theoretical promise and empirical evidence that gave insight into the current landscape of HE ESL. The finding regarding the classification of studies does not contradict the themes but provides context for them revealing what researchers consider as key areas of study and institutions being implemented in the field and why. Therefore, the lack of truly agentic systems among the reviewed literature is not a limitation of the synthesis, but a key finding of the synthesis.

Table 2: Analytical themes

Analytical Theme	The Technological Promise (Prospects)	The Empirical Reality (Pitfalls and Gaps)	Supporting Evidence
1. Adaptive learning and personalisation via agentic AI	Agentic AI promises hyper personalised learning via dynamic profiles from multimodal data (cognitive, behavioural, and motivational), enabling real-time adaptation (Peng and Li, 2025; Holstein, McLaren and Aleven, 2018; Kostopoulos *et al.*, 2025).	Technological asynchrony: Immersive environments are often ‘pedagogically scripted;’ agentic AI for true pedagogical interaction is absent creating a maturity gap (Makransky and Petersen, 2021, p. 951). Evidence scarcity: Short-term, small-scale studies dominate, failing to demonstrate long-term efficacy or skill transfer (Moulieswaran and Prasantha Kumar, 2023, p. 290; Zakaria and Ponniah, 2024).	Differentiated instruction (e.g., targeting B1 speaking) is preferred but immature AI and a lack of longitudinal data are the reality.
2. Immersive technologies in ESL education	VR/AR offers contextual, situated learning that enhances memory and pragmatic skills, reducing anxiety by providing a ‘safe space’	Mixed and null results: Studies find that VR offers no significant advantage over traditional or less immersive methods (Alizadeh and	The promise of reduced anxiety and enhanced skills are counter balanced by equivalent

Analytical Theme	The Technological Promise (Prospects)	The Empirical Reality (Pitfalls and Gaps)	Supporting Evidence
	(Kaplan-Rakowski and Gruber, 2023; Sajja *et al.*, 2024, p. 8 of 23).	Cowie, 2022; Dalgarno and Lee, 2010). Cognitive trade-offs: Immersion may increase cognitive load, causing more errors and longer task times (Bacca-Acosta *et al.*, 2023, p. 98).	outcomes and inefficiencies.
3. Standardisation and assessment (CEFR integration)	CEFR 'can-do' statements provide a framework for AI-driven assessment, balancing standardisation, and individualisation (Ch'Ng *et al.*, 2024, p. 169; North and Piccardo, 2016, p. 455).	Validation gap: Studies rely on self-constructed CEFR mappings lacking empirical validation; claims of progress are often tenuous. Just a few use external raters or standardised tests for triangulation.	The framework exists, but methodological rigour in automated systems is absent.
4. Pedagogical and didactic considerations	The educator's role shifts from knowledge-dispenser to mentor, leveraging AI data for adaptive strategies (Peng and Li, 2025; Luckin *et al.*, 2016, p. 55; Noviandy *et al.*, 2024, p. 81).	Implementation impediment: Barriers include educator resistance, role ambiguity, and insufficient training/support. Readiness for new competencies is not matched by institutional support.	The envisioned shift is pedagogically sound but undermined by a lack of systemic preparation.
5. Ethical and social implications	Data privacy, algorithmic bias mitigation, cultural sensitivity, and equity of access are paramount (Huang, 2023; Kostopoulos *et al.*, 2025; Peng and Li, 2025; Gay, 2018, p. 210).	Persistent risks: Sensitive data use poses privacy risks; biased AI can perpetuate stereotypes; and high VR/AR costs may widen equity gaps.	

4.2 Findings: Review of Empirical Reality and Gaps

4.2.1 Adaptive Learning and Personalisation via Agentic AI

Agentic AI in personalised learning dynamically tailors environments to individual needs, creating a direct connection between content, targeted teaching, and student profiles, including cognitive, social, cultural, and motivational factors (Peng and Li, 2025). Learning analytics frameworks prove that integrating diverse data types enhances understanding and predictive accuracy (Blikstein and Worsley, 2016, p. 220; Papamitsiou and Economides, 2014, p. 49). Through advanced pattern recognition, AI can optimise content for better understanding and retention, providing immediate feedback based on student actions.

Student profile modelling is dynamic, reflecting changes in abilities, preferences, and engagement over time, allowing real-time adjustment of content and activities. Research shows that continuous student profile modelling improves personalisation and learning outcomes (Chrysafiadi and Virvou, 2013, p. 4715; Brusilovsky and Millán, 2007, p. 3). Student profile modelling could aid in combining quantitative data (test scores) with qualitative cues (e.g., tone of voice or attitude shifts). This approach individualises skill development, incorporating contextual data such as time on task and reaction times to optimise learning (Papamitsiou and Economides, 2014, p. 55; Blikstein and Worsley, 2016, p. 225). Real-time analytics reveal peak attention periods, enabling essential content delivery at optimal moments (Holstein *et al.*, 2018). The adaptation loop allows for immediate interpretation and response to student actions, with the system recalibrating based on ongoing data. When aligned with the CEFR, these student profiles become mechanisms for mapping individual learning outcomes to specific descriptor and guide targeted activities (Norshaidatul *et al.*, 2021, p. 64).

Adaptive systems can better assist students with adapting to difficulties and distractions than traditional curriculum designs because they are designed to detect when a student is struggling or losing interest and respond by providing simpler content, additional explanation, or support (Du Plooy, Casteleijn and Franzsen, 2024). The system's ability to be responsive and dynamically adapt is valuable for students with diverse cognitive styles or linguistic backgrounds. It supports more equitable progression by focusing attention and scaffolding in areas where students need it most, rather than limiting them to whole-group pacing. This means that the system does not only offer remedial intervention but also opportunities to progress further in areas of strength. For example, a student excelling in intonation may face higher-level CEFR tasks, while one struggling with listening will receive easier audio until comprehension improves (Norshaidatul *et al.*, 2021, p. 69). As students progress over time, they are assessed and automatically matched to their level of ability, individualising their learning path and providing the student with some autonomy compared to limiting them to progress on a whole group basis.

Adaptive testing platforms use real-time analytics to generate items that accurately measure mastery (Fadieieva, 2023; Du Plooy, Casteleijn and Franzsen, 2024). In group settings, AI-driven moderators assess engagement, encouraging fewer active members and balancing participation, which is a key factor in ESL learning (Sajja *et al.*, 2024). NLP-powered collaborative tools similarly support interactive group participation in online EFL environments, though they remain reactive to student input rather than autonomously directing learning (Toboula, 2023).

Delivering such real-time customisation requires robust architecture and optimisation algorithms capable of millisecond responses (Holstein *et al.*, 2018). AI systems' flexibility could enable rapid experimentation and method testing but it is critical that researchers develop greater methodological rigour, such as using preregistered designs, increased transparency, and have a commitment to open science.

Brusilovsky and Millán (2007, p. 3) have first described how adaptive systems could support the collection of data on student behaviour in a technology mediated environment. Blikstein and Worsley (2016, p. 220) indicate how similar systems could assist student-centred English instruction. These early foundational works provide the basis to progress into AI systems with the potential to capture student behaviour, model it, and use the model to offer personalised instruction. However, their earlier conceptual and methodological contributions developed prior to the rapid evolution of immersive environments. As such, one key challenge is that immersive environments have evolved faster than the AI-based instructional interactions that support them, resulting in a technological asynchrony and immaturity gap.

Immersive environments offer students rich visual opportunities, yet they are still fundamentally pedagogically scripted (Makransky and Petersen, 2021, p. 951). The interplay between a scripted digital immersive world and an interactive adaptive learning platform results in confusion between what is technologically possible and what is pedagogically possible when assessing student achievement, widening the gap between the emerging best practices in ESL and the more traditional, evidence-based practices that have dominated this field since these first ideas developed.

4.2.2 Immersive Technologies in Education

Immersive VR and AR environments allow students to experience a highly interactive environment that combines visual and auditory stimulation with activity-based tasks. These environments enable educators to provide students with tangible representtations of abstract concepts, going beyond what is possible in a traditional classroom. For instance, VR can simulate a scientific experiment, while AR overlays information onto a student’s real-world view. Both technologies prove to increase student engagement and retention of information (Du Plooy, Casteleijn and Franzsen, 2024). Embodied Learning and Situated Cognition theories support this view, indicating that spatial and contextual cues create meaningful connections that directly affect memory formation (Dalgarno and Lee, 2010, p. 20; Makransky and Petersen, 2021, p. 950). This is particularly relevant to ESL instruction, where context plays a significant role in language acquisition; students experience situations rather than merely study them.

AI-created characters in simulated environments can provide students with immediate feedback on their language use, creating a more realistic and low-stakes practice experience (Wik and Hjalmarsson, 2009; Ericsson, Sofkova Hashemi and Lundin, 2023). Wik and Hjalmarsson (2009) describe embodied conversational agents that guide, encourage, and give corrective feedback during role-play simulations, while Ericsson *et al.* (2023) has found that students interacting verbally with virtual humans in simulated everyday scenarios reported the experience as easy, engaging, and safe. In terms of practical application, examples of the use of VR/AR technology in ESL instruction include providing students with live translations and developing VR/AR lesson plans that align with CEFR levels. AI agents can also analyse students’ behaviour, generate new and challenging versions of lessons, and support students as needed. Repeatedly practising conversational responses within a safe environment will help to build student self-confidence and enhance overall communication skills, which are essential in moving through CEFR levels. This can specifically benefit students that feel anxious about speaking English or do not have enough opportunities to interact with native speakers.

The data collected from telemetry devices such as movement sensors, eye tracking devices, and speech recognition systems, can be used to inform AI-driven assessments and evaluations of student performance. As a result, these types of assessments may support accurate and adaptive CEFR-based assessments. However, as Noviandy *et al.* (2024, pp. 85-86) emphasise, with all uses of data collection there is a significant concern about privacy, ethics, and the governance of data.

Critics of immersive technology have expressed concerns about the 'spectacle' of the technology, questioning whether novelty drives engagement rather than genuine learning. Zakaria and Ponniah (2024) have found that when educators collaborate with students to refine VR/AR scenarios, the usability and instructional alignment improve. However, learning gains are still mixed. For example, Muth-mainnah *et al*. (2025) report significant pre-to-post gains in EFL ability using immersive VR-based learning (mean score improving from 58.96 to 84.72, $p < 0.001$), though the study lacks a control group and no comparison with traditional instruction is made. In contrast, Kim, Kim and Cha (2023) find that both metaverse and traditional instruction groups improved within-group, but no significant differences appeared between groups, illustrating the inconsistency of evidence in this area. Most studies are small-scale and short-term, often lacking control groups and underrepresenting diverse students (Moulieswaran and Prasantha Kumar, 2023, p. 290). Longitudinal studies are needed to confirm whether early gains transfer to sustained ability. The lack of standardised protocols also limits cross-study synthesis and meta-analytic integration.

Additionally, immersive learning can introduce heightened sensory engagement and strain, which often increases cognitive load, leading to more errors and longer task durations (Bacca-Acosta *et al.*, 2023, p. 98). This makes it more difficult to track immersive technology's effectiveness during pilot studies, since the observed results may reflect cognitive strain rather than language ability.

Multimodal analysis and eye tracking provide new levels of detail for measuring student engagement during immersive experiences. However, sample sizes remain small, raising concerns about generalisability and replicability.

4.2.3 Standardisation and Assessment (CEFR Integration)

The CEFR provides descriptive criteria for six proficiency levels (A1-C2) with 'can-do' statements that describe what students can accomplish at each level (Council of Europe, 2020). From the reviewed literature, except for Filippone, De Carlo and Di Fuccio (2025) who used Cambridge Assessment-aligned A2/B1 instruments, most studies report language performance gains without any CEFR alignment in measurement. Notably, only one reviewed study uses a validated, CEFR-aligned measurement instrument; CEFR alignment claims in the remaining studies are therefore inferred rather than empirically verified. If proficiency levels are only implied but cannot be mapped in a defensible way, it affects the comparability and interpretive validity of reported outcomes (Ch'Ng *et al.*, 2024; North and Piccardo, 2016). Implementation challenges of CEFR alignment vary significantly across educational contexts, with some nations struggling with educator training and resource allocation (Wok Zaki and Darmi, 2021). The disparity at a system-wide level creates the risk that the validity of assessments conducted by students will be overestimated, thus diminishing the evidence base to support the claim of what is effective.

4.2.4 Pedagogical and Didactic Considerations

In addition to providing a standardised framework (CEFR-aligned) for educators to implement AI-based solutions into their classrooms, the evolving role of the educator as mentor and facilitator will be key to integrating the insights provided by AI systems into their own pedagogical practices (Ravarini, Canavesi and Passerini, 2024). Educators need to continue developing their skills and knowledge so they can effectively evaluate the impact that educational technologies have on their students' learning.

Immersive technology can positively influence social learning by allowing educators to design, co-create, and implement collaborative learning opportunities in multi-user virtual environments, providing students with the opportunity to practise and refine their pragmatic communication skills (Zakaria and Ponniah, 2024). By integrating VR/AR technology with AI tools, educators will have access to automated assessment and intelligent tutoring capabilities, such as speech recognition systems designed to assess student pronunciation. This creates a feedback loop between immersive experiences and the diagnostic assessment of student learning (Holstein *et al.*, 2018; Sajja *et al.*, 2024, p. 8 of 23). Difficulties associated with the use of immersive technology in ESL instruction include the requirement for technical equipment and the need for educator training to effectively integrate VR/AR into teaching practice (Noviandy *et al.*, 2024, p. 87). Policymakers therefore need to invest in programmes that provide educators with continued support. Educator monitoring is advisable, as there is a risk not yet empirically documented in the reviewed literature that overreliance on algorithmically generated feedback may diminish educators' own contextual judgement and reduce their role in providing students with contextually relevant support.

4.2.5 Ethical and Social Implications

Agentic AI and immersive technologies rely on student data to deliver adaptive and personalised feedback increasing the potential risk for misuse of student data and a greater potential for unauthorised access to sensitive information.

Legislative frameworks, such as the General Data Protection Regulation (GDPR) of the European Union, require institutions and companies that collect user data to obtain explicit consent from users before collecting data, to provide users with full transparency regarding the data being collected, how it is being used, how it will be stored and protected, and to implement high standards of security to protect the confidentiality of user data. In terms of the educational applications of AI and immersive technologies in ESL, institutions that deploy these technologies in their classrooms will be required to establish and enforce policies for managing data that meet the requirements of laws such as the GDPR (Huang, 2023; Peng and Li, 2025; Yan *et al.*, 2023). The international scope of language teaching and learning adds another layer of complexity to the regulation of these technologies, as the legal frameworks governing data privacy vary significantly from country to country.

Beyond data privacy concerns, cultural responsiveness must be embedded in AI design to ensure that adaptive systems do not discriminate against diverse students (Ladson-Billings, 2014; Gay, 2018). The deployment of large language models (LLMs) in educational settings raises additional ethical concerns about bias, hallucination of facts, and the need for appropriate human oversight (Yan *et al.*, 2023). AI systems trained on data from specific cultural or linguistic contexts may fail to recognise or appropriately respond to communication patterns from underrepresented groups, thereby reinforcing educational inequities rather than addressing them.

4.3 Future Directions and Innovations in ESL Higher Education

AI and immersive learning environments will continue to evolve, potentially improving natural language processing (NLP) in education due to advancements in pragmatic and affective analysis (Litman, 2016). Emerging research suggests that LLMs may serve as automated evaluators capable of providing consistent, scalable feedback across diverse student populations (Seo *et al.*, 2025). These systems will have to support different channels of communication and develop from recognising patterns to being able to understand semantics and pragmatics and to capture subtle elements of communication such as tone, intent, and social appropriateness (Sajja *et al.*, 2024, p. 23 of 23). This is expected to result in agentic AI developing the ability to engage in conversation with students more naturally and therefore increase the instructional value of pragmatics in ESL.

Future platforms may incorporate other technologies such as facial expression and voice tone recognition in 'plug and play' architectures (Fadieieva, 2023, p. 332). Additionally, automated tutors could be designed to use a student's emotional state to determine specific content and scenarios. Wearable devices equipped with an eye tracker and voice recorder may enable educators and researchers to collect data in real time to use as input for adaptive AI systems that provide immediate feedback (Donnermann, Schaper and Lugrin, 2022, pp. 3-4 of 12). The use of sentiment analysis and multimodal data (text, sound, and images) may create better ways to identify how students are emotionally reacting to the material and may aid in creating diagnostic assessments that evaluate a student's level of proficiency, such as CEFR levels (Sajja *et al.*, 2024, p. 23 of 23).

Algorithms that are adaptive may be able to forecast trends in student skill development and thus enable proactive interventions and provide learning experiences tailored to the individual student using behavioural, cultural, and attitudinal data (Du Plooy, Casteleijn and Franzsen, 2024; Zakaria and Ponniah, 2024). Combining VR/AR with AI could eventually create more seamless and personalised learning experiences for students.

Agentic AI and immersive technologies will need to be carefully integrated into existing educational settings to ensure both effective pedagogy and technology. Contemporary approaches to AI-enhanced course design emphasise modular, flexible architectures that allow educators to customise technological interventions based on specific learning objectives and institutional contexts (Peng and Li, 2025). By utilising modular features and integrating naturally into educators' lesson plans, these AI-enhanced ESL environments can easily be incorporated into existing curricula in either a classroom setting, online, or through some form of blended format (Peng and Li, 2025; Kasztelnik, 2024). For example, VR simulations could allow educators to scale up oral practice in a traditional class by allowing all students to participate at once, while the AI system tracks student activity and connects it to the CEFR assessment systems.

Through distance learning, the AI system can act as a virtual tutor, providing personalised feedback to motivate and engage the student, while also combining the virtual tutor experience with hybrid assessments such as self-tests and group projects that also inform face-to-face instruction (Fadieieva, 2023, p. 339; Holstein *et al.*, 2018). Data collected from both online and physical interactions between the educator and student help to close the gap in a student's progress and align with education frameworks such as CEFR, which links performance data to can-do statements (Ch'Ng *et al.*, 2024, p. 175; Du Plooy, Casteleijn and Franzsen, 2024).

Immersive technologies also allow students to demonstrate their skills in real-world applications and support project-based learning and problem-solving through simulations and interactive AI agents (Fadieieva, 2023, p. 335; Murgatroyd, 2024, p. 6). Overall, HE ESL environments are likely to leverage multimodal, adaptive, and immersive technologies to enhance the educational effectiveness of their programmes.

5. Policy Implications

The primary factors influencing the implementation of agentic AI and immersive technology in HE ESL also include structural challenges, such as infrastructure, equity, policy, and commitment from key stakeholders (Kostopoulos *et al.*, 2025; Peng and Li, 2025). The adoption of these technologies will impact policy at the level of HE in many ways. The review found that immersive environments have outpaced the AI systems meant to support adaptive instruction within them. Funding should prioritise interoperability standards between VR/AR platforms and AI instructional engines rather than the procurement of standalone immersive hardware. Only one of the 42 reviewed studies (Filippone *et al.*, 2025) used a validated CEFR-aligned instrument. Therefore, national education agencies and language testing bodies (e.g., Cambridge Assessment and Alliance Française) should collaborate with technology developers to create open, validated digital assessment instruments benchmarked to CEFR levels. Without this, claims of proficiency gains from AI/VR interventions remain empirically unsubstantiated.

The evidence base is dominated by well-resourced institutions in high-income countries. Mandating equity impact assessments as a condition for educational technology grants and institutions deploying AI-driven learning systems should be required to conduct data protection impact assessments (DPIAs) under the GDPR or equivalent frameworks before collecting multimodal student data. Professional development funding should be ring-fenced and not treated as discretionary to prevent the educator readiness gap from widening.

6. Limitations

Despite the use of a structured and replicable search strategy to adhere to the PRISMA 2020 guidelines, there are several limitations that should be considered when evaluating the findings of this review. The systematic literature review was confined to English language research only. Therefore, potentially relevant non-English studies (particularly those that might be grey literature) may have been excluded due to language bias, and important early-stage findings related to new educational technologies may have been missed in other languages.

The fast-changing field of AI and immersive technologies creates a fundamental temporal limitation for any review of this subject area. Academic publishing is inherently slow compared with the development of technologies; thus, it is possible that many recent developments were not included. As a result, the review represents the state of knowledge in the field at the time of data collection and does not provide a comprehensive picture of all the latest technologies or pedagogies. It should be noted that two references cited in this paper fall outside the standard inclusion window, p. Chiqui Vera *et al.* (2026) published their article after the search cutoff of 15 October 2025 and is therefore included as a post-hoc contextual reference rather than as one of the 42 reviewed studies, while Wik and Hjalmarsson (2009) predates the 2010 lower boundary of the inclusion criteria but is cited as a foundational conceptual reference for embodied conversational agents rather than as part of the systematic synthesis.

The studies included in this review indicate significant variation in terms of their design, technical setup, representative samples, and outcomes measured. As a result, it was impossible to conduct a meta-analysis, and therefore, the results cannot be easily compared or generalised across different contexts. Additionally, many studies had small sample sizes, short-term interventions, and a limited use of standardised measures aligned with CEFR standards (often implied use only), which reduced confidence in making long-term efficacy claims about proficiency gain.

Although quality appraisal based on the CASP tool allowed for a rigorous methodological evaluation of the studies, relying solely on one primary screen tool means that there was a potential source of bias in selecting studies. Although automation tools (such as Elicit) support the initial screening process, decisions made by humans regarding eligibility for inclusion in the review may have biased the final set of studies included.

The thematic analysis was based on the data that were collected in each of the primary studies and were consequently constrained by the same limitations that applied to the original studies, that is, incomplete reporting, poor specification of constructs such as opinions, and poor documentation of the architecture of the AI or design of the immersive technology. Thus, the researcher's conclusions about agentic behaviour, pedagogical alignment, or CEFR integration were often made inferentially rather than being confirmed by the available evidence.

Together, the above limitations reinforce the need for better reporting practices, valid assessment methods, and longitudinal studies in the field of AI mediated language learning. Furthermore, these emphasise the value of regularly revising systematic reviews in fields that evolve quickly, such as AI mediated language learning.

7. Conclusion

Neither the current literature nor the existing infrastructure adequately supports the effective integration of agentic AI and immersive technologies within CEFR-aligned HE ESL. Among the 42 studies reviewed, none implemented a truly agentic system, and only one employed a validated CEFR-aligned measurement instrument. The field frequently mischaracterises adaptive systems as agentic, thereby distorting outcomes and misleading decision-makers. Immersive environments have advanced beyond the AI designed to support them; claims of proficiency gains lack verified measurement; educators are without institutional support; and data governance frameworks are lagging the data that these systems already have collected. These issues represent structural failures rather than incidental ones.

The five themes identified in the review constitute an interconnected system rather than five distinct issues. Addressing any one theme in isolation will not suffice. The convergence of agentic AI and immersive technologies signifies a directional shift for HE ESL. However, the empirical evidence does not substantiate labelling it as a paradigm shift. While the theoretical proposition is coherent and the technological trajectory is evident, the pedagogical and evidential foundations are lacking. Until researchers establish a validated CEFR-aligned measurement base, developers bridge the technology-pedagogy asynchrony, and institutions prioritise educator development and data governance as prerequisites for deployment rather than afterthoughts. As this review has found, agentic AI and immersive technologies in HE

ESL will remain more compelling as a theoretical proposition than as a demonstrated pedagogical reality.

8. Acknowledgements

The author used Paperpal for writing assistance and Elicit AI for supporting the literature review searches during the preparation of this manuscript.

9. Data Availability Statement

All supplementary documents associated with this study are available on the Open Science Framework (OSF) (Kotze, 2025).

10. References

Alizadeh, M. and Cowie, N. (2022) 'The affordances and challenges of virtual reality for language teaching,' *International Journal of TESOL Studies*, 4(3), pp. 5-65. https://doi.org/10.46451/ijts.2022.03.05

Bacca-Acosta, J., Baldiris, S., Fabregat, R. and Ávila, C. (2023) 'Comparative eye-tracking study between a virtual reality system and a desktop environment for learning the prepositions of place in English,' *CALICO Journal*, 40(1), pp. 87-108. https://doi.org/10.1558/cj.22803

Blikstein, P. and Worsley, M. (2016) 'Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks,' *Journal of Learning Analytics*, 3(2), pp. 220-238. https://doi.org/10.18608/jla.2016.32.11

Brusilovsky, P. and Millán, E. (2007) User models for adaptive hypermedia and adaptive educational systems. In Brusilovsky, P., Kobsa, A. and Nejdl, W. eds.: *The Adaptive Web*. Berlin: Springer, pp. 3-53. https://doi.org/10.1007/978-3-540-72079-9_1

CASP (Critical Appraisal Skills Programme). (2018) *CASP checklists*. Oxford: CASP UK.

Chiqui Vera, J.A., Barragan Mejía, E.N., Zambrano Pachay, J.F., Guapacasa Reyes, R.N. and Acosta Barreno, J.K. (2026) 'Enhancing English language skills in higher education through AI: A systematic review of EFL contexts,' *Arandu-UTIC*, 12(4), pp. 3393-3412. https://doi.org/10.69639/arandu.v12i4.1885

Ch'Ng, L.C., Ting, S.H., Sotheeswari, P. and Md Yunus, M. (2024) 'Evaluating students' views on the importance and usefulness of CEFR in speaking test,' *Issues in Language Studies*, 13(1), pp. 166-180. https://doi.org/10.33736/ils.6219.2024

Chrysafiadi, K. and Virvou, M. (2013) 'Student modeling approaches: A literature review for the last decade,' *Expert Systems with Applications*, 40(11), pp. 4715-4729. https://doi.org/10.1016/j.eswa.2013.02.007

Council of Europe. (2020) *Common European Framework of Reference for Languages: Learning, teaching, assessment – companion volume*. Strasbourg: Council of Europe Publishing.

Dalgarno, B. and Lee, M.J.W. (2010) 'What are the learning affordances of 3-D virtual environments?' *British Journal of Educational Technology*, 41(1), pp. 10-32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

Donnermann, M., Schaper, P. and Lugrin, B. (2022) 'Social robots in applied settings: A long-term study on adaptive robotic tutors in higher education,' *Frontiers in Robotics and AI*, 9. 831633. 12 pages. https://doi.org/10.3389/frobt.2022.831633

Du Plooy, E., Casteleijn, D. and Franzsen, D. (2024) 'Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement,' *Heliyon*, 10. e39630. 24 pages. https://doi.org/10.1016/j.heliyon.2024.e39630

Ericsson, E., Sofkova Hashemi, S. and Lundin, J. (2023) 'Fun and frustrating: Students' perspectives on practising speaking English with virtual humans,' *Cogent Education*, 10(1). 2170088. 19 pages. https://doi.org/10.1080/2331186X.2023.2170088

Fadieieva, L.O. (2023) 'Adaptive learning: a cluster-based literaturmodelling (2011-2022),' *Educational Technology Quarterly*, 2023(3), pp. 319-366. https://doi.org/10.55056/etq.613

Filippone, A., De Carlo, M.E. and Di Fuccio, R. (2025) 'Virtual English LAB: The impact of virtual worlds on English language learning and life skills in higher education,' *Excellence and Innovation in Learning and Teaching: Research and Practices*, 10(1), pp. 23-44. https://doi.org/10.3280/exioa1-2025oa20579

Gay, G. (2018) *Culturally responsive teaching: Theory, research, and practice*. 3rd ed. New York: Teachers College Press.

Goyal, S. (2025) 'A critical review of agentic AI: Core technologies, applications, ethical implications, and future research directions,' *Jurnal Masyarakat Informatika*, 16(2), pp. 268-283. https://doi.org/10.14710/jmasif.16.2.77084

Holstein, K., McLaren, B.M. and Aleven, V. (2018) 'Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms.' Paper presented at the 19th International Conference on Artificial Intelligence in Education (AIED 2018), London, pp. 154-168. https://doi.org/10.1007/978-3-319-93843-1_12

Huang, L. (2023) 'Ethics of artificial intelligence in education: Student privacy and data protection,' *Science Insights Education Frontiers*, 20, pp. 297-311. https://doi.org/10.15354/sief.23.re202

Kaplan-Rakowski, R. and Gruber, A. (2023) 'The impact of high-immersion virtual reality on foreign language anxiety,' *Smart Learning Environments*, 10. 46. 18 pages. https://doi.org/10.1186/s40561-023-00263-9

Kasztelnik, K. (2024) 'Artificial intelligence-assisted curriculum development: Innovations in designing educational content for the 21st century learner,' *Journal of Higher Education Theory and Practice*, 24(11), pp. 51-59. https://doi.org/10.33423/jhetp.v24i11.7367

Kim, H.S., Kim, N.Y. and Cha, Y. (2023) 'Exploring the potential of metaverse as a future learning platform for enhancing EFL learners' English proficiency,' *Korean Journal of English Language and Linguistics*, 23, pp. 220-236. https://doi.org/10.15738/kjell.23.202303.220

Kostopoulos, G., Gkamas, V., Rigou, M. and Kotsiantis, S. (2025) 'Agentic AI in education: State of the art and future directions,' *IEEE Access*, 13, pp. 177467-177491. https://doi.org/10.1109/ACCESS.2025.3620473

Kotze, C. (2025) *Integrating agentic AI and immersive technologies into standardised and adaptive CEFR-aligned ESL learning environments in higher education: A systematic review of prospects and pitfalls*. OSF. 9 December 2025. Available at: https://osf.io/b75mj. (Accessed 9 December 2025).

Ladson-Billings, G. (2014) ‘Culturally relevant pedagogy 2.0: A.k.a. the remix,’ *Harvard Educational Review*, 84(1), pp. 74-84. https://doi.org/10.17763/haer.84.1.p2rj131485484751

Lin, G.Y., Jhang, C.C. and Wang, Y.S. (2024) ‘Factors affecting parental intention to use AI-based social robots for children’s ESL learning,’ *Education and Information Technologies*, 29(6), pp. 6059-6086. https://doi.org/10.1007/s10639-023-12023-w

Litman, D. (2016) ‘Natural language processing for enhancing teaching and learning,’ *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), pp. 4170-4176. https://doi.org/10.1609/aaai.v30i1.9879

Luckin, R., Holmes, W., Griffiths, M. and Forcier, L.B. (2016) *Intelligence unleashed: An argument for AI in education*. London: Pearson.

Makransky, G. and Petersen, G.B. (2021) ‘The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality,’ *Educational Psychology Review*, 33(3), pp. 937-958. https://doi.org/10.1007/s10648-020-09586-2

Moulieswaran, N. and Prasantha Kumar, N.S. (2023) ‘Investigating ESL learners’ perception and problem towards artificial intelligence (AI)-assisted English language learning and teaching,’ *World Journal of English Language*, 13(5), pp. 290-302. https://doi.org/10.5430/wjel.v13n5p290

Murgatroyd, S. (2024) ‘Artificial intelligence and future of higher education,’ *Revista Paraguaya de Educación a Distancia*, 5(1), pp. 4-11. https://doi.org/10.56152/reped2024-vol5num1-art1

Muthmainnah, M., Cardoso, L., Marzuki, A.G. and Al Yakin, A. (2025) ‘A new innovative metaverse ecosystem: VR-based human interaction enhances EFL learners’ transferable skills,’ *Discover Sustainability*, 2025(6). 156. n.p. https://doi.org/10.1007/s43621-025-00913-7

North, B. and Piccardo, E. (2016) ‘Developing illustrative descriptors of aspects of mediation for the Common European Framework of Reference (CEFR),’ *Language Teaching*, 49(3), pp. 455-459. https://doi.org/10.1017/S0261444816000100

Norshaidatul, M.N., Siti Hajar, H., Melor, M.Y. and Harwati, H. (2021) ‘CEFR for languages and its effective implementation in secondary schools in Malaysia,’ *Asian Journal of Assessment in Teaching and Learning*, 11(1), pp. 63-72. https://doi.org/10.37134/ajatel.vol11.1.6.2021

Noviandy, T.R., Maulana, I., Ramadhan, G.A.P., Rizki, J. and Sofyan, H. (2024) ‘Embrace, do not avoid: Reimagining higher education with generative artificial intelligence,’ *Journal of Educational Management and Learning*, 2(2), pp. 81-91. https://doi.org/10.60084/jeml.v2i2.233

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P. and Moher, D. (2021) ‘The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,’ *BMJ*, 372, n71. 9 pages. https://doi.org/10.1136/bmj.n71

Papamitsiou, Z. and Economides, A.A. (2014) ‘Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence,’ *Educational Technology and Society*, 17(4), pp. 49-64.

Peng, J. and Li, Y. (2025) ‘Frontiers of artificial intelligence for personalized learning in higher education: A systematic review of leading articles,’ *Applied Sciences*, 15(18). 10096. 31 pages. https://doi.org/10.3390/app151810096

Ravarini, A., Canavesi, A. and Passerini, K. (2024) ‘From users to allies: Exploring educator and generative AI roles in shaping the future of higher education.’ Paper presented at the 10th International Conference on Higher Education Advances (HEAd’24), Valencia, Spain, pp. 312-319. https://doi.org/10.4995/HEAd24.2024.17345

Sajja, R., Sermet, Y., Cwiertny, D. and Demir, I. (2024) ‘Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education,’ *Information*, 15(10). 596. 23 pages. https://doi.org/10.3390/info15100596

Seo, H., Hwang, T., Jung, J., Kang, H., Namgoong, H., Lee, Y. and Jung, S. (2025) ‘Large language models as evaluators in education: Verification of feedback consistency and accuracy,’ *Applied Sciences*, 15(2). 671. 35 pages. https://doi.org/10.3390/app15020671

Sweller, J., Van Merriënboer, J.J.G. and Paas, F. (2019) ‘Cognitive architecture and instructional design: 20 years later,’ *Educational Psychology Review*, 31(2), pp. 261-292. https://doi.org/10.1007/s10648-019-09465-5

Thomas, J. and Harden, A. (2008) ‘Methods for the thematic synthesis of qualitative research in systematic reviews,’ *BMC Medical Research Methodology*, 8(1). 45. 10 pages. https://doi.org/10.1186/1471-2288-8-45

Wang, D. and Huang, X. (2025) ‘Transforming education through artificial intelligence and immersive technologies: Enhancing learning experiences,’ *Interactive Learning Environments*, 33(7), pp. 4546-4565. https://doi.org/10.1080/10494820.2025.2465451

Wik, P. and Hjalmarsson, A. (2009) ‘Embodied conversational agents in computer assisted language learning,’ *Speech Communication*, 51(10), pp. 1024-1037. https://doi.org/10.1016/j.specom.2009.05.006

Wok Zaki, A. and Darmi, R. (2021) ‘The implementation of CEFR in ESL learning: Why does it matter to the Malaysian education system?’ *Asian Journal of Assessment in Teaching and Learning*, 11(2), pp. 1-13. https://doi.org/10.37134/ajatel.vol11.2.1.2021

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y. and Gašević, D. (2023) ‘Practical and ethical challenges of large language models in education: A systematic scoping review,’ *British Journal of Educational Technology*, 55(1), pp. 90-112. https://doi.org/10.1111/bjet.13370

Yu, D. (2023) ‘AI-empowered metaverse learning simulation technology application.’ Paper presented at the 2023 IEEE International Conference on Metaverse (iMETA). 6 pages. https://doi.org/10.1109/iMETA59369.2023.10294830

Zakaria, N.Y.K. and Ponniah, T. (2024) ‘Analysing the impact of artificial intelligence in ESL education: A systematic review,’ *International Journal of Academic Research in Progressive Education and Development*, 13(4), pp. 2649-2668. https://doi.org/10.6007/ijarped/v13-i4/23895

Appendix A: Supplementary Material

Figure 1. PRISMA 2020 flow diagram for study selection

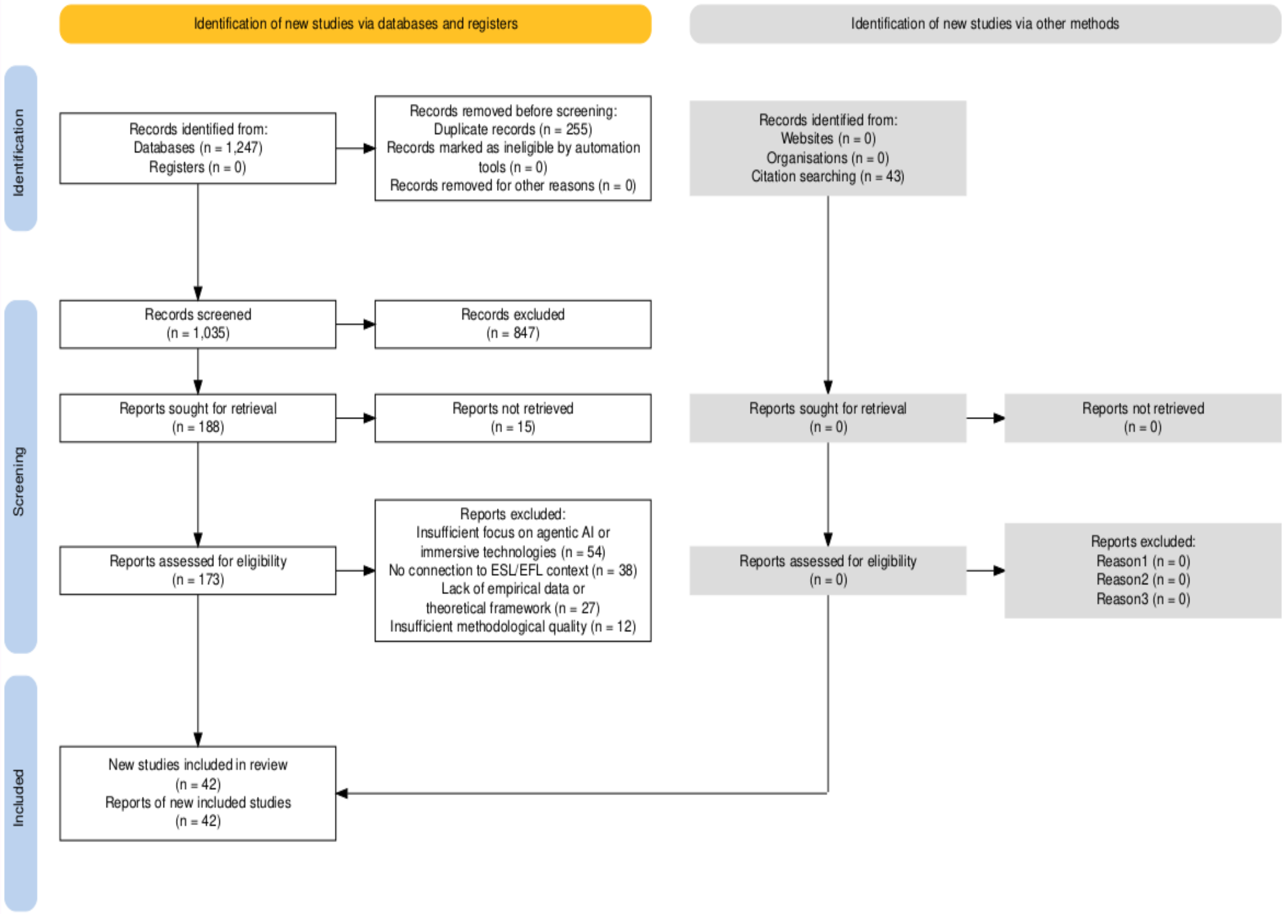

Figure 2: Quality appraisal diagram

Quality Category	Number of Studies	Percentage	Score Range
High Quality	26	61.9%	8.0-10.0
Medium Quality	14	33.3%	5.0-7.5
Lower Quality	2	4.8%	3.0-4.5
Total	**42**	**100%**	**3.0-10.0**

Table 3: AI system classification against agentic criteria

Criteria key: A = Autonomy | G = Goal-directed behaviour | D = Dynamic decision-making | P = Proactivity

Symbol key: ✓ = met | χ = not met | ∂ = partially met

Study	A / G / D / P	Classification	Rationale
Bacca-Acosta *et al.* (2023)	χ χ χ χ	Not classified (immersive)	VR eye-tracking only; no AI instructional component
Ericsson *et al.* (2023)	χ χ ∂ χ	Rule-based / Scripted	Virtual humans follow scripts; no instructional adaptation
Filippone *et al.* (2025)	χ χ χ χ	Not classified (immersive)	Frame VR environment; no AI adaptive component

Study	A / G / D / P	Classification	Rationale
Holstein *et al.* (2018)	∂ ✗ ✓ ✓	ITS (proactive features)	Proactively surfaces student data; educator retains final decisions
Noviandy *et al.* (2024)	✗ ✗ ∂ ✗	Adaptive	Generative AI discussed; no autonomous goal setting
Rodriguez and Hemachandran (2023)	∂ ✗ ∂ ✗	Adaptive	AI avatars for engagement; no goal setting
Sajja *et al.* (2024)	∂ ∂ ✓ ∂	ITS (approaching agentic)	Adaptive assistant with proactive features; within structured domain
Sarwat *et al.* (2024)	✗ ✗ ✗ ✗	Not classified (immersive)	Immersive reader for comprehension; no AI adaptation
Seo *et al.* (2025)	✗ ✗ ∂ ✗	Adaptive	LLM feedback provider; no autonomous instructional decisions
Song (2024)	✗ ✗ ∂ ✗	Adaptive	Modular AI course design; parameter-based
Toboula (2023)	✗ ✗ ∂ ✗	Adaptive	NLP collaboration tools; reactive to input
Wik and Hjalmarsson (2009)	✗ ✗ ∂ ✗	Rule-based / Scripted	Conversational scripts with limited branching

Note. Studies included in this table met the inclusion criteria for AI-supported instructional systems. Conceptual, review-based, and non-AI studies were excluded from classification. A system was classified as agentic only when all four criteria (autonomy, goal-directed behaviour, dynamic decision-making, and proactivity) were simultaneously present. Classification was conducted independently by two reviewers, with disagreements resolved through discussion.