International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Review Article)

Check for updates

Leveraging Artificial Intelligence for Simulation and Visualisation in STEM Education: A Theory-Informed Narrative Synthesis

Dinesh Deckker [1,*] and Subhashini Sumanasekara [2]

[1] *Faculty of Arts, Science and Technology, Wrexham University, United Kingdom.*
[2] *Faculty of Computing and Social Sciences, University of Gloucestershire, United Kingdom.*

International Journal of Science and Research Archive, 2025, 16(02), 901-913

Publication history: Received on 10 July 2025; revised on 17 August 2025; accepted on 19 August 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.16.2.2412

Abstract

This review synthesises emerging evidence (2022–2025) on the role of artificial intelligence (AI) in enhancing simulation and visualisation within STEM education. Traditional instructional approaches often face challenges in representing abstract concepts, ensuring equitable access to laboratory experiences, and fostering higher-order cognitive skills. AI-driven simulations address these gaps by enabling dynamic interaction, personalised feedback, and real-time adaptation of task complexity, grounded in constructivist and socio-constructivist learning theories. Drawing on recent empirical studies, this article examines how generative AI, multi-agent platforms, reinforcement learning controllers, and explainable AI frameworks transform conceptual understanding, engagement, self-regulated learning, and model-based reasoning. Findings highlight consistent gains in conceptual grasp, motivation, and misconception repair, particularly when AI systems are designed with inquiry cycles, cognitive load principles, and teacher oversight. However, persistent challenges remain in equity, scalability, data governance, and ethical safeguards, with gaps in early education contexts and longitudinal research. The paper proposes a structured research agenda that emphasises inclusive design, cross-disciplinary collaboration, and transparency. By integrating theoretical foundations with technological advancements, this synthesis contributes a framework for the effective, equitable, and accountable application of AI-driven simulation and visualisation in STEM teaching and learning.

Keywords: Artificial Intelligence; STEM Education; Simulation; Visualisation; Pedagogy; Ethics

1. Introduction

STEM education remains central to preparing learners for an increasingly complex, technology-driven world, yet traditional instructional approaches often struggle to convey abstract concepts and replicate authentic problem-solving environments (Kefalis et al., 2025). Laboratory experiments, while valuable, can be limited by cost, safety, and accessibility, particularly in resource-constrained contexts. Moreover, passive learning modes are insufficient for developing the higher-order cognitive skills and adaptive expertise demanded by contemporary STEM professions (León, 2025).

Artificial intelligence (AI)-powered simulation and visualisation tools offer a paradigm shift by enabling learners to interact dynamically with complex systems, manipulate variables in real time, and visualise invisible processes such as molecular interactions or electromagnetic fields (Marquez-Carpintero et al., 2023). These systems extend the principles of constructivist learning (Piaget, 1954; Vygotsky, 1978; Bruner, 1966) by allowing learners to iteratively explore, hypothesise, and test within safe, adaptive, and data-rich environments. Unlike static digital models, AI-driven simulations can personalise feedback, identify misconceptions, and adjust task complexity, thereby aligning with the learner's zone of proximal development (Vygotsky, 1978; El Fathi et al., 2025).

* Corresponding author: Dinesh Deckker

This article provides a critical, theory-informed analysis of how AI-enhanced simulation and visualisation tools are transforming STEM education. It synthesises recent empirical evidence (2022–2025) to:

- Examine the pedagogical underpinnings of AI-supported simulation;
- Evaluate their capacity to improve conceptual understanding, learner engagement, and skill acquisition;
- Identify persistent challenges, including equity, ethics, and scalability; and
- Outline future research pathways. By integrating foundational educational theory with cutting-edge applications, this paper aims to contribute a scholarly framework for the effective and ethical use of AI-driven simulation and visualisation in STEM teaching and learning.

1.1. Methodology

This study employs a narrative review methodology to critically examine and synthesise recent scholarship (2022–2025) on the use of AI-enhanced simulation and visualisation tools in STEM education. A narrative review was selected as it enables the integration of emerging, multi-disciplinary evidence with established pedagogical theories, supporting both descriptive mapping and analytical interpretation (Green et al., 2006). This approach is particularly suited to rapidly developing fields where technology and pedagogy evolve in parallel.

1.2. Literature Search Strategy

A targeted search was undertaken in Scopus and Web of Science in March 2025 using Boolean combinations of keywords including *"AI in STEM education"*, *"simulation AND AI"*, *"visualisation tools AND AI"*, *"adaptive learning systems"*, and *"multimodal AI tutors"*. Additional studies were identified through backwards and forward citation tracking of key articles.

1.2.1. Inclusion criteria

- Peer-reviewed journal articles published between 2022 and 2025.
- Empirical studies examining AI-driven simulation or visualisation in STEM contexts.
- Studies reporting learning outcomes or pedagogical analysis.
- Explicit theoretical grounding or articulation of learning mechanisms.

1.2.2. Exclusion criteria

- Non-peer-reviewed publications or grey literature.
- Studies outside STEM education.
- Purely conceptual discussions without empirical evidence.

1.3. Data Extraction and Organisation

From each selected study, details were extracted on:

- Theoretical frameworks referenced (e.g., Piaget's constructivism, Vygotsky's socio-constructivism, Bruner's cognitive apprenticeship, Mayer's multimedia learning principles).
- AI technology features (e.g., adaptive scaffolds, generative visual metaphors, multi-agent simulation roles).
- Reported learning outcomes (e.g., conceptual understanding, metacognitive skill development, learner engagement).
- Study design, participant profile, and educational setting.
- The extracted data were organised into construct–feature–measure matrices and pedagogical mechanism maps to trace the relationships between AI affordances, cognitive and social learning processes, and educational outcomes.

The review applies an interpretive, theory-informed lens, connecting recent empirical findings to foundational learning theories. Emphasis is placed on identifying mechanistic pathways—how specific AI features (e.g., personalised visualisation, adaptive feedback loops) activate cognitive, metacognitive, and social processes that contribute to STEM learning gains.

1.4. Limitations

As a narrative review, the search strategy was selective rather than exhaustive, prioritising conceptual richness and methodological diversity over completeness. The exclusion of non-English publications may have omitted relevant work, and variability in outcome reporting precludes quantitative synthesis. Nonetheless, the chosen approach facilitates a depth-oriented, integrative analysis that bridges theoretical foundations with emerging practice.

2. Theoretical & Pedagogical Foundations

AI-enhanced simulation and visualisation tools align most directly with constructivist and socio-constructivist traditions by enabling iterative hypothesis-testing, dialogic support, and contingent scaffolding. However, recent syntheses note that the explicit theorisation of such tools is uneven; many studies emphasise implementation and outcomes while making only cursory links to learning theory (Lee et al., 2024; Wang et al., 2024; León, 2025). This section, therefore, articulates how major theoretical lenses should inform the design and evaluation of AI-mediated simulations, and where contemporary work falls short.

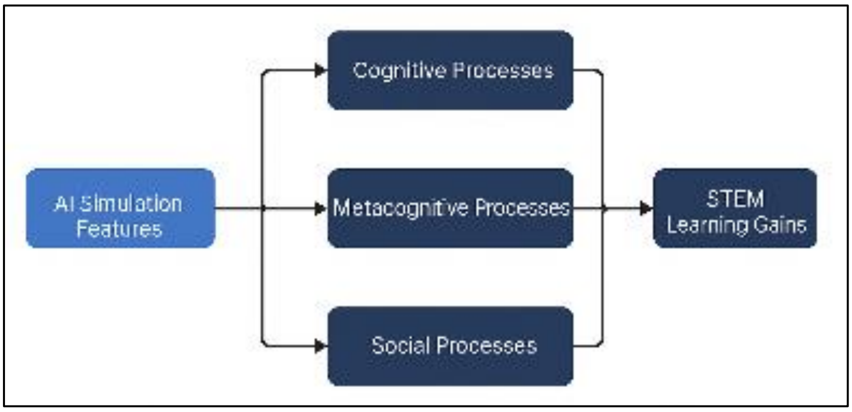

Figure 1 Pedagogical Mechanism Map Linking AI Simulation Features to Cognitive, Metacognitive, and Social Processes in STEM Learning

2.1. Constructivism, Inquiry, and Adaptive Guidance

Constructivism predicts learning gains when students manipulate variables, receive formative cues, and iteratively refine mental models. AI-driven simulations operationalise these mechanisms by diagnosing misconceptions in real time and tuning task complexity to the learner's state. A quasi-experimental study integrating ChatGPT through a constructivist inquiry-based prompting framework demonstrated improved conceptual understanding and misconception repair in first-year engineering thermodynamics, exemplifying theory-consistent adaptive feedback (El Fathi et al., 2025). Systematic reviews of digital simulations similarly report consistent benefits for conceptual grasp and problem-solving when interactivity and feedback are foregrounded (Kefalis et al., 2025). However, across K-12 AI in education reviews, theoretical commitments are often implicit rather than designed in, complicating the accumulation of evidence about why effects arise (Lee et al., 2024; Wang et al., 2024).

Future studies should pre-register theory-linked design features (e.g., hypothesis cycles, fading scaffolds, error-contingent prompts) and test them as mechanisms, not just interface affordances (León, 2025).

2.2. Socio-constructivism and Teacher–AI Orchestration

Socio-constructivist models emphasise mediated learning within the zone of proximal development (ZPD). Recent work proposes didactic models for integrated STEM that embed collaborative inquiry and structured guidance, offering a blueprint for lesson design into which AI tutors or agents can be orchestrated as *secondary* mediators rather than replacements for teachers (Toma et al., 2024). Reviews of AI in STEM ecosystems argue for transdisciplinary communication frameworks so that educators, designers, and policy actors co-specify roles for human and machine scaffolds, including accountability boundaries and escalation rules (León, 2025). Empirically, teacher readiness and AI literacy remain binding constraints on effective orchestration; reviews identify variability in teacher confidence and in schools' capacity to sustain data-driven workflows (Lee et al., 2024; Wang et al., 2024).

Studies should model division of cognitive labour across teacher, peer, and AI, and measure orchestration quality (e.g., timing of AI interventions, granularity of hints) as predictors of learning.

2.3. Cognitive Load, Visualisation, and Model-Based Reasoning

Simulation and visualisation efficacy is bounded by cognitive architecture. Contemporary advances in multimedia learning and cognitive load research specify design principles—signalling, segmentation, spatial contiguity—that are directly actionable for AI systems that generate or adapt visuals on the fly (Mayer, 2024; Çeken & Taşkın, 2022). Experimental and design-based studies in 2023–2025 show that small changes to realism, shape distinctness, and signalling can materially alter load and learning from complex visualisations (Skulmowski, 2023; Deibl et al., 2023; Bali et al., 2025). Augmented/virtual reality work indicates that applied CTML/CLT-aligned choices reduce overload and improve outcomes in immersive settings (Candido et al., 2025; Jian et al., 2024). Meta-analytic evidence on concept mapping—a visual externalisation of knowledge—corroborates medium effects on STEM achievement, suggesting that AI-assisted concept-map generation can serve as a principled scaffold for model-based reasoning (Wang et al., 2025).

Researchers should report load-sensitive telemetry (e.g., interaction density, visual complexity metrics) and couple it with outcomes, enabling AI systems to learn design policies (e.g., when to simplify a field line plot or segment a process animation).

2.4. Where the Theory is Thin: A Programmatic Gap

Across recent systematic reviews, explicit theory-driven mechanism testing is rare relative to outcome comparisons; theoretical constructs are referenced but not operationalised (Lee et al., 2024; Wang et al., 2024; León, 2025). Few studies factorially manipulate theoretically meaningful parameters (e.g., fixed vs. adaptive scaffolding; peer-first vs. AI-first mediation). This limits transferability and masks which ingredients matter in AI simulations.

A next-generation research agenda should adopt theory-to-design matrices (construct → feature → measure), report manipulation checks for intended scaffolds and publish reusable design rationales to accelerate cumulative science.

---

3. Technological Landscape & Design Architectures

AI-enhanced simulation and visualisation in STEM now coalesces around four architectural lines: (i) generative/MLLM pipelines that produce adaptive visuals and task scaffolds; (ii) multi-agent environments that stage social/experiential practice; (iii) policy-learning controllers (reinforcement learning and related optimisation) embedded in simulators; and (iv) explainability-first toolchains that expose model rationale to learners and teachers. Design choices across these lines should be evaluated as mechanisms (what they do pedagogically), not merely as features (what they look like).

3.1. Generative/MLLM Pipelines for Dynamic Visualisation and Scaffolding

Modern systems couple large language models (LLMs) and multimodal LLMs (MLLMs) to: (a) translate learner prompts or misconceptions into targeted visuals (e.g., field-line sketches, molecule dynamics, parametric plots); (b) regulate task difficulty by re-phrasing goals and constraints; and (c) generate worked examples and counter-examples aligned to multimedia design principles. Two design anchors follow:

Load-aware visual design. Empirical multimedia research specifies load-sensitive principles—signalling, segmenting, and spatial contiguity—that directly inform AI-generated visual scaffolds; MLLM renderers should implement these as first-class constraints rather than after-the-fact aesthetics (Mayer, 2024).

Capability bounds and affordances. Generative AI can boost feedback, content generation, and regulation of learning, but reliability and alignment concerns remain; thus, visual outputs require guardrails (e.g., canonical references, unit checks, provenance labels) and teacher-in-the-loop validation (Giannakos et al., 2024).

Mechanistic hypothesis. When MLLMs generate contrastive visual metaphors (correct vs. common misconception side-by-side) under CTML constraints, they should reduce extraneous load and sharpen model-based reasoning; studies should report load proxies and transfer outcomes to test this claim (Mayer, 2024).

3.2. Multi-Agent Simulation Environments (MASE)

Multi-agent architectures instantiate social and experiential practice: AI agents assume roles (peer, evaluator, client) to create repeatable, feedback-rich scenarios (e.g., lab troubleshooting, design reviews). Two converging strands of evidence support this line:

Educational agents for practice and SRL. Controlled studies with AI-powered virtual humans show gains in self-regulated learning behaviours among STEM learners, indicating that agentic practice can shape planning/monitoring processes that generalise beyond the simulator (Glick et al., 2024).

LLM-driven multi-agent platforms. Peer-reviewed computer-science venues now document toolchains where LLMs govern agent behaviour and interaction protocols at scale—both in domain-general swarm settings and education-specific classroom simulations (NAACL 2025). These reports detail role templates, interaction controllers, and reflection loops that can be translated to STEM scenarios (Jimenez-Romero et al., 2025; Zhang et al., 2025).

MASE should specify (i) division of cognitive labour (teacher/peer/AI roles), (ii) feedback topology (who critiques whom, when, and how), and (iii) assessment hooks (rubrics aligned to the scenario's epistemic aims). Reporting these as independent variables enables mechanism testing rather than one-off demos (Zhang et al., 2025).

3.3. Policy-Learning Controllers in Simulators

Reinforcement learning (RL) and related optimisation are increasingly embedded to adapt task flow (progression, hint timing) and system dynamics (e.g., stability thresholds in physics engines). A recent systematic review shows RL's growing use for personalised interventions and adaptive sequencing, while calling for stronger educational outcome measures and transparency about reward design (Riedmann et al., 2025).

Controllers should be educationally shaped: reward functions must encode epistemic goals (e.g., conceptual change, transfer), not mere completion speed; policies should be auditable and pedagogically legible to teachers.

3.4. Explainability-First Simulation Toolchains

For simulations to be trusted and instructionally usable, the system must expose why a recommendation, trajectory, or visual state occurred. Education-specific XAI frameworks distinguish the needs of learners (formative, stepwise, conceptual) from those of teachers (diagnostic, summative, cohort-level), offering concrete design levers (e.g., counterfactual explanations, feature attribution overlays, provenance traces) (Khosravi et al., 2022).

XAI should be embedded into the interaction loop: when a simulation adjusts a parameter set or proposes a new path, it should simultaneously present an explanation suitable for the audience (learner vs. instructor), with toggles for depth and uncertainty disclosure (Khosravi et al., 2022).

3.5. Cross-cutting Implementation Notes

Usability and acceptance. Student experience studies in STEM show positive uptake of LLM tools but surface anxieties about reliability and assessment alignment; these should be treated as design constraints (Valeri et al., 2025).

Transdisciplinary orchestration. Reviews in STEM-education contexts emphasise communication frameworks that align designer, teacher, and policy roles—critical for scaling labs and classrooms that mix physical apparatus, simulations, and AI tutoring (León, 2025).

4. Pedagogical Impact Analysis

4.1. Learning outcomes and conceptual understanding

Across recent evidence, interactive simulations consistently improve conceptual understanding and problem-solving when designs emphasise manipulation, immediate feedback, and inquiry cycles. A 2025 systematic review of 31 empirical studies reports positive effects on core STEM outcomes and learner engagement, with the most substantial gains under inquiry-based conditions and when feedback is tightly coupled to task actions (Kefalis et al., 2025).

Generative-AI-supported tutoring appears effective for misconception repair under challenging topics. In an engineering thermodynamics course, integrating ChatGPT within a constructivist, inquiry-based workflow led to significant pre-post gains on targeted concepts, with parallel improvements in acceptance and engagement tracked longitudinally (El Fathi et al., 2025). Mechanistically, benefits arose from rapid clarification prompts and adaptive re-explanations aligned to students' error patterns.

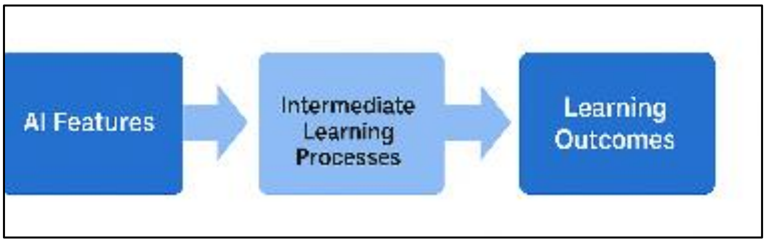

Figure 2 Effect Chain Diagram Mapping AI Features to Learning Processes and Outcomes in STEM Education

4.2. Higher-order thinking, metacognition, and self-regulation

AI-mediated practice can develop self-regulated learning (SRL) behaviours that generalise beyond a single task. In a controlled study with STEM majors, AI-powered virtual human modules improved planning and self-monitoring tool use relative to controls, indicating that agent-led micro-interventions can shift SRL routines (Glick, 2024).

Generative systems may also function as "objects-to-think-with," supporting reflective reasoning and creative problem framing in STEM contexts. A 2023 peer-reviewed case study found that Chats provoked productive reflection and conceptual articulation, while also surfacing reliability limits that require teacher mediation—applicable for designing safer simulation prompts and verification steps (Vasconcelos & dos Santos, 2023).

4.3. Engagement and motivation in immersive visual environments

Immersive visualisation (AR/VR) typically boosts motivation and performance compared with traditional media, provided cognitive-load controls (signalling, segmentation, spatial contiguity) are respected. A 2023 review in *Heliyon* synthesised substantial gains in engagement and task performance with AR/VR, while cautioning about distraction and load when designs neglect multimedia principles (Al-Ansi et al., 2023). Complementary work demonstrates that applying cognitive theory of multimedia learning (CTML) to immersive/extended reality lessons reduces overload and improves outcomes—design guidance directly transferable to AI-generated visualisations (Mayer, 2024).

4.4. Comparative effectiveness and boundary conditions

Comparative findings indicate AI-augmented simulations outperform static or non-adaptive versions when (i) feedback is immediate and specific, (ii) tasks are progressively adapted, and (iii) visual complexity is tuned to learner state. The 2025 review (Kefalis et al., 2025) highlights more potent effects in inquiry-rich settings; however, primary and special-education phases remain under-represented, limiting generalisability claims (research gap). In addition, studies emphasise trust and usability as moderators of impact: students report benefits but also anxieties regarding accuracy and assessment alignment, signalling the need for provenance cues and teacher-in-the-loop verification in classroom rollouts (Valeri et al., 2025).

4.5. Durability, transfer, and assessment

Longitudinal evidence remains thin. While thermodynamics findings suggest durable gains over a semester (El Fathi et al., 2025), few studies test far-transfer or delayed retention with robust assessment designs. Future work should employ multi-timepoint designs and construct-aligned instruments (e.g., conceptual inventories plus transfer tasks), and report mechanism-level telemetry (e.g., hint timing, revision cycles, visual-complexity metrics) to link design moves to learning change.

Table 1 Pedagogical Impact of AI-Enhanced Simulation and Visualisation Tools in STEM Education (2022–2025)

Construct	Metric(s)	Reported Effect	Key Design Features	Citation
Conceptual Understanding	Pre-/post-test conceptual scores; misconception analysis	Significant gains in understanding and problem-solving, particularly in inquiry-based contexts	Interactive variable manipulation, immediate feedback, embedded inquiry cycles	Kefalis et al. (2025)

Misconception Repair	Targeted diagnostic items; long-term retention scores	Reduction of persistent misconceptions; maintained gains over a semester	ChatGPT integration with constructivist inquiry prompts; adaptive re-explanations	El Fathi et al., 2025(2025)
Self-Regulated Learning (SRL)	SRL behaviour inventories; log analysis of planning/monitoring actions	Increased planning and self-monitoring behaviours	AI-powered virtual-human modules delivering micro-interventions	Glick (2024)
Reflective Reasoning & Creativity	Qualitative coding of dialogue; creativity rubrics	Enhanced reflective articulation and creative framing of problems	Generative AI (ChatGPT, Bing Chat) as "objects-to-think-with"; prompts encouraging conceptual reflection	Vasconcelos & dos Santos (2023)
Engagement & Motivation (Immersive)	Self-report engagement scales; performance tasks	Higher engagement and task performance in immersive settings	AR/VR with cognitive-load–controlled visual design (signalling, segmentation)	Al-Ansi et al. (2023)
Cognitive Load Management	NASA-TLX; dual-task performance	Lower extraneous load and improved learning when applying CTML principles	Visual signalling, segmentation, and spatial contiguity in AI-generated visuals	Mayer (2024)
Trust & Usability	Likert-scale trust/usability surveys; qualitative feedback	Positive perceptions of usefulness, tempered by reliability concerns	Provenance cues; teacher-in-the-loop validation of AI outputs	Valeri et al. (2025)

5. Equity, Ethics & Trust Considerations

AI-enhanced simulation and visualisation can widen access to powerful STEM experiences. Yet, they also risk amplifying inequities and eroding trust if fairness, privacy, and transparency are not designed in from the outset. A recent transdisciplinary review of AI in STEM education argues that ethical questions are not peripheral but *constitutive* of effective implementation. It proposes a Transdisciplinary Communication (TDC) framework to align designers, educators, and policy actors around shared norms for safety, accountability, and inclusivity (León, 2025).

5.1. Bias and Fairness in AI-Mediated Simulations

Bias may enter through skewed training data, opaque optimisation targets, or differential interface demands. A mapping study on machine-learning fairness in higher education highlights gaps in how bias is defined, measured, and mitigated in educational software pipelines, and calls for task-specific fairness metrics and audits integrated into development workflows (Pham et al., 2025). For simulation tools, fairness checks should be linked to pedagogic tasks (e.g., item difficulty adaptation, feedback timing) rather than generic parity metrics, with mitigation reported through pre-registered audit plans and error margins.

5.2. Privacy, Data Protection, and Governance

Simulation platforms often combine learner traces (clickstreams, speech, gaze) with assessment outcomes. A systematic review of privacy and data protection in learning analytics identifies eight persistent, intertwined risks across the analytics lifecycle and recommends privacy-by-design and evidence-based mitigations rather than ad-hoc consent forms (Liu et al., 2023). In line with statutory regimes such as FERPA and GDPR (U.S. Department of Education, n.d.), designers should adopt data minimisation and purpose limitation, provide provenance receipts for model updates, and expose user-facing controls for retention and sharing.

5.3. Transparency and Explainability for Learners and Teachers

Trustworthy classroom deployment requires auditable explanations tuned to audience needs. An education-specific explainable AI framework outlines six dimensions of explainability (e.g., user goals, timing, modality). It demonstrates how explanations should differ for learners (formative, step-wise) versus teachers (diagnostic, cohort-level) (Khosravi et al., 2022). Embedding such explanations within the interaction loop supports accountability without overwhelming users. For example, pairing an adaptive change in a physics simulation with a contextual explanation of what changed, why, and with what level of confidence can improve transparency and trust.

5.4. Trust, Acceptance, and Human Oversight

Trust predicts uptake and sustained use. A recent study validated a trust instrument for AI-powered educational technologies, showing that perceived usefulness and transparency are necessary but insufficient; control over the AI's behaviour (e.g., override, pause, reveal reasoning) significantly shapes user attitudes (Nazaretsky et al., 2025). Cross-national survey evidence further indicates that teachers' trust is sensitive to prior experience, perceived benefits, and institutional safeguards, reinforcing the value of teacher-in-the-loop controls in classroom simulations (Viberg et al., 2024).

5.5. Ethical Risk Identification and Continuous Safeguards

A systematic review of AI in K-12 education highlights recurring ethical risks—bias propagation, opacity, and safety—and recommends developer-side improvements in data reliability and model fairness, alongside teacher-side review of outputs for potential harms (Zhu et al., 2025). This underscores the need for joint responsibility across vendors and schools, supported by continuous post-deployment monitoring such as fairness drift checks and incident reporting.

5.6. Equity and Accessibility

Even well-designed systems can exacerbate inequities if access is unaffordable or if interfaces assume high bandwidth, English-dominant text, or specific motor/visual abilities. Ethical frameworks for AI in STEM call for inclusive design baselines—offline modes, low-spec fallbacks, screen-reader compatibility, and localisation—and for impact reporting disaggregated by demographic and contextual variables (León, 2025).

6. Implementation & Scalability Challenges

Although AI-driven simulation and visualisation tools offer compelling pedagogical affordances, their adoption in STEM education is constrained by implementation and scalability barriers. These span teacher readiness, infrastructural capacity, and systemic inequities, each of which can determine whether innovations remain isolated pilots or become sustainable, institution-wide practices.

6.1. Usability and System Acceptance

Usability directly shapes both student and teacher acceptance of AI-supported simulations. In a 2025 study on integrating generative AI into thermodynamics instruction, usability perceptions—particularly ease of navigation, clarity of feedback, and alignment with course objectives emerged as key predictors of sustained student engagement (El Fathi et al., 2025). When usability is low, even high-quality adaptive feedback mechanisms are underutilised, reducing return on investment. This finding aligns with broader human–computer interaction education research, which identifies perceived usefulness and ease of use as dominant acceptance factors for emerging learning technologies (Valeri et al., 2025).

6.2. Teacher Digital Literacy and Professional Development

Scaling AI-enhanced simulation requires teachers to possess not only subject knowledge but also digital literacy sufficient to configure, monitor, and troubleshoot AI systems. Cross-national surveys show considerable variability in teacher preparedness for AI integration, with disparities linked to access to professional development, institutional support, and prior experience with educational technology (Viberg et al., 2024). Without targeted upskilling, teachers may under-utilise adaptive features or fail to interpret AI-generated analytics, undermining learning gains.

6.3. Institutional Capacity and Infrastructure

The deployment of AI-driven simulation platforms can impose substantial infrastructure demands, including high-performance computing resources, reliable broadband, and device compatibility. In low-resource contexts, these demands exacerbate digital divides and restrict equitable participation (León, 2025). Even in well-resourced institutions, integration may require upgrades to legacy systems, data privacy compliance measures, and robust IT support for scaling beyond experimental phases.

6.4. Resource Inequities and Equity of Access

Institutional resource inequities, including variation in technology budgets, technical support, and access to modern hardware, can result in uneven adoption. Ethical frameworks for AI in STEM education stress the importance of inclusive procurement policies, shared infrastructure models, and low-spec or offline-capable versions of simulation tools to mitigate inequity (León, 2025).

6.5. Organisational Change and Policy Alignment

Sustained scaling also depends on policy alignment at institutional and governmental levels. This includes clear guidelines for data governance, curriculum integration, and evaluation metrics. Evidence from AI-in-education policy studies suggests that adoption is more successful when initiatives are embedded in a broader digital transformation strategy rather than pursued as isolated technology deployments (Nazaretsky et al., 2025).

Table 2 Implementation & Scalability Challenges of AI-Enhanced Simulation and Visualisation in STEM Education (2022–2025)

Construct	Metric(s) / Evidence Source	Reported Effect / Challenge	Key Factors / Design Implications	Citation
Usability & Student Acceptance	Student surveys (SUS, TAM), engagement analytics	Usability perceptions (navigation ease, feedback clarity, curricular alignment) predict sustained engagement.	Prioritise intuitive interfaces, embedded help, and straightforward task–outcome mapping.	El Fathi et al. (2025)
Teacher Digital Literacy	Cross-national teacher surveys; digital skills self-assessments	Significant variation in AI preparedness linked to PD access and prior EdTech experience.	Invest in targeted PD on AI tools, analytics interpretation, and troubleshooting.	Viberg et al. (2024)
Institutional Capacity	Infrastructure audits; IT support logs	High computational, bandwidth, and compatibility demands hinder scaling.	Plan phased infrastructure upgrades; integrate with existing LMS/IT systems.	León (2025)
Resource Inequities	Budget analysis; access gap studies	Low-resource contexts face exclusion due to hardware and connectivity gaps.	Develop low spec/offline modes; pool resources via regional consortia	León (2025)
Policy & Organisational Alignment	Policy reviews; institutional strategy documents	Adoption is stronger when part of a digital transformation agenda, not isolated pilots.	Align procurement, curriculum, and governance under clear AI integration policies.	Nazaretsky et al. (2025)
Trust & Long-term Adoption	Longitudinal usage data; teacher/student trust instruments	Sustained use depends on transparency, oversight, and user control features.	Embed teacher-in-the-loop oversight; publish system capability/failure mode documentation.	Nazaretsky et al. (2025)

7. Future Research & Innovation Pathways

AI-enhanced simulation and visualisation in STEM education is advancing rapidly, but the current literature leaves several conceptual, methodological, and equity-related gaps that warrant a structured research agenda.

7.1. Integration with Emerging Technologies

The next phase of innovation will likely involve synergistic integration between AI-driven simulations and other immersive technologies, including extended reality (XR), augmented reality (AR), virtual reality (VR), and intelligent laboratories. Studies in immersive STEM environments show that XR can significantly increase engagement and comprehension when coupled with cognitive-load–aware design (Al-Ansi et al., 2023; Candido et al., 2025). Embedding

AI into these platforms could enable adaptive XR simulations, where the complexity of the virtual scene, the sequencing of tasks, and the scaffolding provided are dynamically adjusted to learner profiles. Intelligent labs, integrating AI-controlled apparatus with real-time analytics, could bridge the gap between *in silico* and hands-on experimentation, supporting hybrid pedagogical models that blend constructivist exploration (Piaget, 1954; Vygotsky, 1978) with data-driven adaptivity.

7.2. Addressing Methodological Gaps

Current research is skewed towards higher education, with primary and special education contexts markedly under-represented. A systematic review of digital simulations in STEM education found that while secondary and tertiary levels dominate empirical studies, early learning stages and diverse learner populations (e.g., neurodiverse students) receive limited attention (Kefalis et al., 2025). Expanding research to these groups would help validate whether the same adaptive feedback loops and multimodal visualisations are effective, or whether tailored designs are necessary. Moreover, few studies adopt experimental designs that isolate causal mechanisms, such as factorial experiments varying scaffold type, AI involvement level, and visual complexity. Greater use of design-based research (DBR) approaches could also help iterate tools in authentic classroom settings while maintaining theoretical fidelity.

7.3. Longitudinal and Equity-Focused Studies

Evidence on long-term learning retention, transfer, and equity impacts remains sparse. Although some semester-long studies report durable conceptual gains (El Fathi et al., 2025), there is a need for multi-timepoint evaluations extending across academic years to assess whether AI-supported simulation advantages persist without continued exposure. In parallel, research must explicitly track equity outcomes, disaggregating performance and engagement data by socio-economic status, gender, language background, and disability status (León, 2025). Without such disaggregation, potential biases in adaptive feedback or visual representation may remain undetected. Equity-centred AI research should integrate fairness auditing methods from machine-learning scholarship (Pham et al., 2025) directly into study protocols, allowing bias detection during, not after, deployment.

7.4. Building Cross-Disciplinary Research Consortia

To address the complex interplay between pedagogy, technology, and ethics, cross-disciplinary consortia—linking education researchers, AI developers, cognitive scientists, and policy specialists—are essential. Such collaborations could develop shared evaluation frameworks that capture not only learning gains but also cognitive load, trust, usability, and fairness, enabling meta-analysis across contexts. Funding agencies and policy bodies should prioritise these collaborative structures to accelerate cumulative, generalisable knowledge.

8. Conclusion

The evidence reviewed in this study demonstrates that AI-driven simulation and visualisation tools in STEM education operate through multiple, complementary pedagogical mechanisms. Generative systems—extensive and multimodal language models enable adaptive content creation, generate metaphorical and contrastive visualisations, and offer personalised scaffolds aligned to constructivist and cognitive-load principles (Marquez-Carpintero et al., 2023; Mayer, 2024). Multi-agent platforms simulate complex social and collaborative problem-solving environments, providing iterative practice in authentic contexts and fostering self-regulated learning behaviours (Glick, 2024; Zhang et al., 2025). Interactive simulation environments, when designed with real-time feedback, adjustable complexity, and learner-controlled exploration, enhance conceptual understanding, engagement, and transfer (Kefalis et al., 2025; El Fathi et al., 2025). Collectively, these technologies extend the reach of foundational learning theories (Piaget, 1954; Vygotsky, 1978; Bruner, 1966) into data-rich, adaptive environments that can be continuously optimised through empirical feedback.

For educators, the findings highlight the importance of aligning AI-enabled simulations with curricular goals, embedding opportunities for reflection, and maintaining teacher oversight to ensure contextual accuracy and pedagogical appropriateness. Professional development should address not only tool operation but also interpretation of AI-generated analytics and integration into assessment practices (Viberg et al., 2024). For developers, the priority is to design AI systems that are usable, transparent, and ethically robust, embedding explainability features, provenance tracking, and fairness auditing into the design architecture (Khosravi et al., 2022; Pham et al., 2025). Both stakeholders should collaborate to ensure that adaptive algorithms are tuned not merely for engagement, but for durable learning and equitable access.

This review advances the literature by synthesising pedagogical theory, technological capability, and ethical considerations into a theory-informed, evidence-based framework for AI-supported STEM simulation and visualisation.

It addresses methodological gaps by identifying under-researched contexts (e.g., primary and special education), equity challenges, and the need for longitudinal impact studies. By integrating insights from cognitive science, socio-constructivist pedagogy, and AI ethics, the work provides a roadmap for future research and development that bridges the divide between innovative design and educational practice. In doing so, it contributes to a growing scholarly discourse on how AI can transform STEM education in ways that are effective, inclusive, and accountable.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Al-Ansi, A. M., Jaboob, M., Garad, A., & Al-Ansi, A. (2023). *Analyzing augmented reality (AR) and virtual reality (VR)* recent development in education. Social Sciences & Humanities Open, 8(1), Article 100532. https://doi.org/10.1016/j.ssaho.2023.100532

[2] Bali, C., Várkonyi, G., Szabó, M., & Zsidó, A. N. (2025). The impact of visual cues on reducing cognitive load in interactive storybooks for children. Journal of Experimental Child Psychology, 260, 106320. Advance online publication. https://doi.org/10.1016/j.jecp.2025.106320

[3] Bruner, J. S. (1966). Toward a theory of instruction. Harvard University Press.

[4] Candido, V., & Cattaneo, A. (2025). Applying cognitive theory of multimedia learning principles to augmented reality and its effects on cognitive load and learning outcomes. Computers in Human Behavior Reports, 18, Article 100678. Advance online publication. https://doi.org/10.1016/j.chbr.2025.100678

[5] Çeken, B., & Taşkın, N. (2022). Multimedia learning principles in different learning environments: A systematic review. Smart Learning Environments, 9, 19. https://doi.org/10.1186/s40561-022-00200-2

[6] Deibl, I., Zumbach, J., & Fleischer, T. (2023). Visualization and metacognitive scaffolding in learning from animations. Social Sciences & Humanities Open, 8(1), Article 100601. https://doi.org/10.1016/j.ssaho.2023.100601

[7] El Fathi, T., Saad, A., Larhzil, H., Lamri, D., & Al Ibrahmi, E. M. (2025). Integrating generative AI into STEM education: Enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance. Disciplinary and Interdisciplinary Science Education Research, 7, Article 6. https://doi.org/10.1186/s43031-025-00125-z

[8] Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2024). The promise and challenges of generative AI in education. Behaviour & Information Technology, 44(11), 2518–2544. Advance online publication. https://doi.org/10.1080/0144929X.2024.2394886

[9] Glick, D., Miedijensky, S., & Zhang, H. (2024). Examining the effect of AI-powered virtual-human training on STEM majors' self-regulated learning behavior. Frontiers in Education, 9, Article 1465207. https://doi.org/10.3389/feduc.2024.1465207

[10] Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. Journal of Chiropractic Medicine, 5(3), 101–117. https://doi.org/10.1016/S0899-3467(07)60142-6

[11] Jian, Y., Abu Bakar, J.A. Comparing cognitive load in learning spatial ability: immersive learning environment vs. digital learning media. Discov Sustain 5, 111 (2024). https://doi.org/10.1007/s43621-024-00310-6

[12] Jimenez-Romero, C., Yegenoglu, A., & Blum, C. (2025). Multi-agent systems powered by large language models: Applications in swarm intelligence. Frontiers in Artificial Intelligence, 8, 1593017. https://doi.org/10.3389/frai.2025.1593017

[13] Kefalis, C., Skordoulis, C., & Drigas, A. (2025). Digital simulations in STEM education: Insights from recent empirical studies – A systematic review. Encyclopedia, 5(1), 10. https://doi.org/10.3390/encyclopedia5010010

[14] Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. Computers & Education: Artificial Intelligence, 3, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074

[15] Lee, S. J., & Kwon, K. (2024). A systematic review of AI education in K-12 classrooms from 2018 to 2023: Topics, strategies, and learning outcomes. Computers & Education: Artificial Intelligence, 6, Article 100211. https://doi.org/10.1016/j.caeai.2024.100211

[16] Leon, C., Lipuma, J., & Oviedo-Torres, X. (2025). Artificial intelligence in STEM education: A transdisciplinary framework for engagement and innovation. Frontiers in Education, 10, Article 1619888. https://doi.org/10.3389/feduc.2025.1619888Liu, Q., Chen, S., & Viberg, O. (2023). Understanding privacy and data protection issues in learning analytics: A systematic review. British Journal of Educational Technology, 54(6), 2531–2554. https://doi.org/10.1111/bjet.13388

[17] Marquez-Carpintero, L., Viejo, D., & Cazorla, M. (2025). Enhancing engineering and STEM education with vision and multimodal large language models to predict student attention. IEEE Access, 13, 114681–114695. https://doi.org/10.1109/ACCESS.2025.3584025

[18] Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. Educational Psychology Review, 36, 8. https://doi.org/10.1007/s10648-023-09842-1

[19] Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2025). The critical role of trust in adopting AI-powered educational technology for learning: An instrument for measuring student perceptions. Computers & Education: Artificial Intelligence, 8, Article 100368. https://doi.org/10.1016/j.caeai.2025.100368

[20] Pham, N., Pham Ngoc, H., & Nguyen-Duc, A. (2025). Fairness for machine learning software in education: A systematic mapping study. Journal of Systems and Software, 219, 112244. https://doi.org/10.1016/j.jss.2024.112244

[21] Piaget, J. (1954). The construction of reality in the child. Basic Books.

[22] Riedmann, A., Schaper, P. & Lugrin, B. Reinforcement Learning in Education: A Systematic Literature Review. Int J Artif Intell Educ (2025). https://doi.org/10.1007/s40593-025-00494-6

[23] Skulmowski, A. (2023). Shape distinctness and segmentation benefit learning from realistic visualizations, while dimensionality and perspective play a minor role. Computers & Education: X Reality, 2, 100015. https://doi.org/10.1016/j.cexr.2023.100015

[24] Toma, R., Yánez-Pérez, I. & Meneses-Villagrá, J. Towards a Socio-Constructivist Didactic Model for Integrated STEM Education. Interchange 55, 75–91 (2024). https://doi.org/10.1007/s10780-024-09513-2

[25] U.S. Department of Education, Student Privacy Policy Office. (n.d.). Family Educational Rights and Privacy Act (FERPA). In Protecting Student Privacy. Retrieved July 17, 2025, from https://studentprivacy.ed.gov/ferpa

[26] Valeri, F., Nilsson, P., & Cederqvist, A.-M. (2025). Exploring students' experience of ChatGPT in STEM education. Computers & Education: Artificial Intelligence, 8, Article 100360. https://doi.org/10.1016/j.caeai.2024.100360

[27] Vasconcelos, M. A. R., & dos Santos, R. P. (2023). Enhancing STEM learning with ChatGPT and Bing Chat as objects-to-think-with: A case study. Eurasia Journal of Mathematics, Science and Technology Education, 19(7), em2296. https://doi.org/10.29333/ejmste/13313

[28] Viberg, O., Cukurova, M., Feldman-Maggor, Y. et al. What Explains Teachers' Trust in AI in Education Across Six Countries?. Int J Artif Intell Educ (2024). https://doi.org/10.1007/s40593-024-00433-x

[29] Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

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

[31] Wang, X.-M., Wang, J.-L., Xu, S.-Y., & Xu, S.-J. (2025). Concept mapping in STEM education: A meta-analysis of its impact on students' achievement (2004–2023). International Journal of STEM Education, 12, Article 30. https://doi.org/10.1186/s40594-025-00554-2

[32] Zhang, Z., Zhang-Li, D., Yu, J., Gong, L., Zhou, J., Hao, Z., Jiang, J., Cao, J., Liu, H., Liu, Z., Hou, L., & Li, J. (2025, April). Simulating classroom education with LLM-empowered agents. In L. Chiruzzo, A. Ritter, & L. Wang (Eds.), Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational

Linguistics: Human Language Technologies (Vol. 1: Long Papers, pp. 10364–10379). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.naacl-long.520

[33] Zhu, H., Sun, Y., & Yang, J. (2025). Towards responsible artificial intelligence in education: A systematic review on identifying and mitigating ethical risks. Humanities and Social Sciences Communications, 12, Article 1111. https://doi.org/10.1057/s41599-025-05252-6