Artificial Intelligence Education Studies
Volume 1, Issue 2, 2025, Page 1–18
https://doi.org/10.6914/aiese.010201

Artificial Intelligence Education Studies

Applications of Large Multimodal Models (LMMs) in STEM Education: From Visual Explanations to Virtual Experiments

Changkui LI

Hong Kong Integration Research Institute, Email:lichangkui@gmail.com, https://orcid.org/0000-0001-7446-0198

Abstract

Generative Artificial Intelligence (GAI) refers to a class of AI systems capable of creating novel, coherent, and contextually relevant content—such as text, images, audio, and video—based on patterns learned from extensive training datasets. The public release and rapid refinement of large language models (LLMs) like ChatGPT have accelerated the adoption of GAI across various medical specialties, offering new tools for education, clinical simulation, and research. Dermatology training, which heavily relies on visual pattern recognition and requires extensive exposure to diverse morphological presentations, faces persistent challenges such as uneven distribution of educational resources, limited patient exposure for rare conditions, and variability in teaching quality. Exploring the integration of GAI into pedagogical frameworks offers innovative approaches to address these challenges, potentially enhancing the quality, standardization, scalability, and accessibility of dermatology education. This comprehensive review examines the core concepts and technical foundations of GAI, highlights its specific applications within dermatology teaching and learning—including simulated case generation, personalized learning pathways, and academic support—and discusses the current limitations, practical challenges, and ethical considerations surrounding its use. The aim is to provide a balanced perspective on the significant potential of GAI for transforming dermatology education and to offer evidence-based insights to guide future exploration, implementation, and policy development.

Keywords Large Multimodal Models (LMMs); STEM Education; Visual Explanations; Virtual Laboratories / Virtual Experiments; Critical AI Literacy

1 Introduction

1.1 The Emergence of Large Multimodal Models (LMMs) in the Educational Landscape

The rapid evolution of artificial intelligence (AI) has culminated in the development of Large Multimodal Models (LMMs), sophisticated systems capable of processing and integrating information from a variety of data types.[1] This marks a significant departure from earlier Large Language Models (LLMs), which primarily operated within the textual domain. The capacity of LMMs to understand and generate content across modalities such as images, audio, video, and even programming code holds profound implications for numerous fields, with education standing as a prime beneficiary. Within the educational sphere, Science, Technology, Engineering, and Mathematics (STEM) disciplines are uniquely positioned to leverage these multimodal capabilities. STEM fields inherently rely on diverse forms of representation, including complex diagrams, mathematical equations, experimental data, and physical phenomena, all of which can be more holistically addressed by LMMs than by their unimodal counterparts.

The advent of LMMs signals a potential transformation in educational methodologies, promising to move beyond predominantly text-based digital interactions towards richer, more contextualized, and interactive learning experiences.[1] This transition is not merely an incremental technological improvement; rather, it represents a potential paradigm shift. Where LLMs could assist with tasks like summarizing texts or answering factual questions, LMMs

© The Authors 2025. Article History Received: April 19, 2025; Accpted: May 10, 2025; Published: June 30, 2025. ***Artificial Intelligence Education Studies***, ISSN 3079-8086 (print), ISSN 3079-8094 (online), a Quarterly, founded on 2025, published by Creative Publishing Co., Limited. Email: wtocom@gmail.com, https://ai-es.org, https://cpcl.hk.

can engage with the core visual, auditory, and interactive elements that are fundamental to STEM understanding. This capability allows AI to function less as a text-based assistant and more as a co-constructor of multimodal comprehension, fundamentally altering how AI can be integrated into the learning process.

1.2 Defining LMMs: Capabilities Beyond Text for STEM

LMMs are AI models engineered to process, understand, and generate information from multiple types of data concurrently.[1] Unlike LLMs, which are confined to textual input and output, LMMs can interpret and create content involving text, images, audio, video, and, increasingly, more specialized data forms like programming code or sensor readings relevant to scientific experiments.[4] Prominent examples of LMMs with significant multimodal processing capabilities relevant to STEM include Google's Gemini, which can reason across text, images, video, audio, and code[4]; Meta's Llama 3.2 Vision, which excels at image-text tasks such as chart and diagram understanding[5]; NVIDIA's NVLM family, designed for robust multimodal reasoning and high-resolution image handling[5]; and OpenAI's GPT-4o, which demonstrates advanced capabilities in scientific problem-solving involving visual data.[6]

The core distinction between LMMs and LLMs lies in this expanded data processing spectrum. For STEM education, this means LMMs can engage with scientific diagrams, analyze video demonstrations of experiments, generate visual representations of abstract concepts, or even produce code for simulations—tasks that are largely outside the scope of text-only LLMs.[2] This enhanced capacity allows LMMs to support a wider and deeper range of pedagogical activities in STEM.

The following table provides a comparative overview of LMMs and LLMs in the context of educational applications:

Table 1: Comparative Overview of LMMs and LLMs for Educational Applications

Feature	Large Language Models (LLMs)	Large Multimodal Models (LMMs)
Data Modalities	Primarily text-based[2]	Text, images, audio, video, code, sensor data[1]
Integration Capabilities	Limited to text; cannot inherently combine with other data types	Strong at combining and understanding diverse data types simultaneously[2]
Typical Applications	Writing assistance, translation, Q&A, summarization[2]	Image captioning, visual Q&A, video analysis, text-to-image/video generation, multimodal data analysis[2]
STEM-Specific Potential	Explaining concepts textually, generating problem sets, basic coding	Generating/interpreting diagrams, visualizing data, creating simulations, analyzing experimental videos, interactive tutorials
Key Limitations	Inability to process visual or auditory STEM content directly	Technical complexity, data demands, potential for cross-modal hallucination, alignment of heterogeneous data[2]

1.3 Rationale and Objectives for LMMs in STEM Education

The rationale for integrating LMMs into STEM education is compelling. Many core STEM concepts are inherently abstract and benefit significantly from visual, interactive, and experimental modes of engagement, which LMMs are uniquely equipped to support.[8] Traditional teaching methods, and even earlier forms of educational technology, often struggle to provide sufficiently rich and adaptable multimodal experiences at scale. LMMs, with their capacity to generate and interpret diverse data formats, offer the potential to overcome these limitations. For instance, the creation of high-quality interactive simulations or dynamic visual explanations has historically been resource-intensive.[11] LMMs could potentially automate or semi-automate this process[13], thereby enabling pedagogical designs, such as highly personalized inquiry-driven virtual labs, that were previously too complex or costly to implement widely.

This paper aims to critically review the current and potential applications of LMMs in STEM education, with a specific focus on two pivotal areas: the enhancement of visual explanations and the facilitation of virtual experiments. It will analyze the associated pedagogical frameworks necessary for effective integration, explore innovative assessment strategies suitable for LMM-rich learning environments, and scrutinize the ethical considerations that accompany such powerful technologies. A further objective is to delineate promising future research directions that can guide the responsible and impactful development and deployment of LMMs in STEM education. The introduction of LMMs may also necessitate a re-evaluation of "literacy" in STEM. Beyond traditional textual and numerical literacy, students and educators will increasingly require skills in multimodal AI interaction—the ability to effectively prompt, interpret, and critically evaluate LMM outputs across various data types. This emerging form of literacy will be crucial for navigating an educational and professional landscape increasingly influenced by advanced AI.

2 Theoretical Underpinnings for LMM Integration in STEM

The effective integration of LMMs into STEM education necessitates a robust theoretical foundation that considers how students learn with and through these complex technologies. Established learning theories provide a starting point, though they may require adaptation, while new pedagogical frameworks are emerging to specifically address the unique affordances and challenges of multimodal AI.

2.1 Relevant Learning Theories in the Age of LMMs

Several learning theories are particularly relevant to the integration of LMMs. First, **Constructivism** posits that learners actively construct their own understanding and knowledge of the world through experiencing things and reflecting on those experiences. LMMs align well with constructivist principles by providing tools and environments where students can actively engage with multimodal content and simulations.[15] For example, an LMM could generate a dynamic model of a cell, allowing students to manipulate variables and observe outcomes, thereby constructing their understanding of cellular processes. The capacity of LMMs to generate personalized learning environments and diverse resources can further facilitate this exploration and discovery, empowering students to build knowledge in a manner tailored to their individual pathways. Second, **Cognitive Load Theory (CLT)** is concerned with the limitations of working memory during learning.[17] Instructional design should aim to optimize cognitive load by minimizing extraneous load (imposed by poor design), managing intrinsic load (inherent to the material's complexity), and fostering germane load (devoted to learning and schema construction). LMMs present a dual potential here. On one hand, they can reduce extraneous cognitive load by providing clear, concise visual explanations or by simplifying complex information into more digestible multimodal formats.[17] On the other hand, poorly designed LMM interactions, inaccurate outputs, or overwhelming amounts of generated information could inadvertently increase cognitive load, hindering learning.[18] A study by Zhai et al. highlighted that while LLMs might reduce mental effort (a component of cognitive load), this could come at the cost of compromised depth in student scientific inquiry.[18] This "cognitive ease at a cost" suggests that the interaction with LMMs must be carefully managed. Traditional CLT did not anticipate AI agency or the dynamic nature of AI-generated content; thus, CLT requires re-evaluation. The cognitive load associated with verifying AI outputs or managing the interaction itself introduces new variables that must be considered in instructional design. Finally, **Mayer's Multimedia Learning Principles** offer research-based guidelines for designing effective multimedia instruction, emphasizing that people learn better from words and pictures than from words alone.[20] Key principles include the Coherence Principle (excluding extraneous material), the Signaling Principle (highlighting essential material), the Redundancy Principle (avoiding presenting identical information in multiple formats simultaneously if one is narration), and the Spatial and Temporal Contiguity Principles (placing corresponding words and pictures near each other and presenting them simultaneously).[20] These principles are highly relevant for designing and evaluating LMM-generated visual and auditory explanations. For LMM-generated content to be pedagogically effective, it must adhere to these guidelines to manage cognitive load and promote meaningful learning. However, since LMMs generate dynamic multimedia, these principles need to be applied not just to the final output, but to the process of generation and interaction. For instance, an LMM that generates an overly complex diagram violating the Coherence Principle, or an animation with poorly timed narration violating the Temporal Contiguity Principle, could impede rather than support learning.

2.2 Pedagogical Frameworks for Multimodal AI in STEM

As LMMs become more integrated into education, specific pedagogical frameworks are needed to guide their effective and ethical use in STEM. One such approach, the **Multidimensional Frameworks for GenAI in Education** outlined by Educause, identifies definitional, systemic, cognitive processing, and pedagogical dimensions for viewing GenAI.[24] The pedagogical dimension is particularly relevant, outlining developmental levels from naïve use through AI competence and AI literacy to ethical use. This provides a broad structure for understanding LMM adoption. Another key framework is **Inquiry-Based Learning (IBL)**, which emphasizes student-centered learning where students ask questions and investigate phenomena.[25] LMMs can significantly support IBL by providing tools for hypothesis generation and experimentation in virtual environments, such as in the PrimaryAI curriculum which integrates AI concepts into an IBL framework.[26] The **ARCHED (AI for Responsible, Collaborative, Human-centered Education Instructional Design)** framework proposes a structured workflow emphasizing collaboration between educators and AI, with educators as the primary decision-makers.[27] Lastly, **Tseng & Warschauer's Framework**, originally for AI writing tools, offers valuable principles for general AI literacy in STEM: Understand, Access, Prompt, Corroborate, and Incorporate.[28] An updated version also emphasizes "think first" and "reflect" on AI use.[30]

Effective pedagogical frameworks for LMMs will likely need to be inherently meta-cognitive. They must explicitly teach students about the AI's underlying mechanisms, its potential biases, how its "knowledge" is constructed, and how to engage with it as a powerful but fallible partner. Frameworks like ARCHED[27] which emphasize human agency, and Tseng & Warschauer's[31] which includes corroboration, point towards this need. The focus on epistemic insights[32] further underscores that pedagogy must move beyond using LMMs as black boxes and involve teaching students about the LMMs themselves.

Table 2: Key Pedagogical Frameworks for LMM Integration in STEM Education

Framework Name	Core Principles	Key LMM Affordances Leveraged	STEM Application Examples
Multidimensional GenAI Framework	Defines levels of AI proficiency (Naïve Use, Competence, Literacy, Ethics); considers definitional, systemic, cognitive, pedagogical dimensions.	Content generation, personalized feedback, data analysis.	Developing institutional AI policies, designing AI literacy curricula, guiding ethical AI use in research projects.
Inquiry-Based Learning (IBL) + AI	Student-led questioning, investigation, experimentation, and conclusion-drawing, supported by AI tools.	Simulation generation, data visualization, hypothesis testing support, access to information.	Students use LMMs to design virtual experiments, analyze simulated data, or explore complex scientific questions (e.g., PrimaryAI).
ARCHED Framework	Human-AI collaboration in instructional design; human agency maintained; transparency; Bloom's taxonomy as foundation.	Generation of pedagogical options, alignment checking with learning objectives.	Educators use LMMs to brainstorm lesson activities, generate visual aids, or draft assessment items, with final curation and refinement by the educator.
Tseng & Warschauer's Framework (adapted for STEM)	Understand AI, Access tools, Prompt effectively, Corroborate outputs, Incorporate thoughtfully; Think first, Reflect after.	Information retrieval, content generation (text, visuals, code), problem-solving assistance.	Students learn to critically use LMMs for research, generating initial drafts of lab reports, or debugging code for simulations.
Human-Centric AI-First (HCAIF) Framework	Attribution of AI use, student reflection on AI's role, personalized learning, continuous feedback, competency development.	Personalized content/exercise creation, feedback generation, assessment support.	Students use LMMs for research and content creation, clearly attributing AI contributions and reflecting on the learning process in journals.

2.3 Cultivating Critical AI Literacy and Epistemic Vigilance

Beyond understanding how to operate LMMs, students in STEM must develop critical AI literacy. This involves not just knowing how to use these tools, but how to think critically about their outputs, their inherent limitations, and their ethical implications.[34] A core component of this is fostering "epistemic vigilance"—an attitude of healthy skepticism and a commitment to verifying information generated by LMMs, especially in scientific contexts where accuracy is paramount.

Students need to grasp the epistemic nature of AI-generated knowledge: how LMMs "know" what they know (i.e., through pattern recognition in vast datasets, not through genuine understanding or empirical validation in the scientific sense), the sources and certainty of this "knowledge," and its inherent limitations.[32] This includes understanding AI's specific applications in scientific endeavors and recognizing both the similarities (e.g., use of data, pattern identification) and crucial differences (e.g., lack of causal reasoning, potential for ungrounded "hallucinations") in the epistemological approaches of science versus AI.[32] The development of critical AI literacy is therefore not merely an auxiliary skill but an ethical imperative in STEM. Uncritical acceptance of LMM-generated outputs, whether textual, visual, or simulative, could lead to the propagation of misinformation, the reinforcement of flawed scientific models, or the adoption of biased conclusions, ultimately undermining the integrity of scientific learning and practice.

3 LMMs for Enhanced Visual Explanations in STEM Disciplines

Visualizations are indispensable in STEM, serving to clarify complex phenomena, represent abstract concepts, and communicate data. LMMs offer novel capabilities in both generating and interpreting scientific visuals, potentially transforming how students engage with and understand these critical representations.

3.1 Generating and Interpreting Complex Scientific Visualizations

LMMs are demonstrating increasing proficiency in creating novel visual representations of scientific concepts, data, and processes.[13] This goes beyond simple image generation to include the creation of structured diagrams and interpretable visual outputs.

A notable development is the "Visualizing Thought" framework, where LMMs generate conceptual diagrams—for instance, by producing executable Matplotlib code—to assist in their own reasoning and planning processes.[13] This approach allows an LMM to visually "show its work," making its intermediate reasoning steps more transparent. In planning tasks like the Blocksworld domain, this method significantly improved GPT-4o's accuracy from 35.5% to 90.2%.[13] This ability of LMMs to generate code for visualizations is a more profound capability than merely producing static images. It opens pathways for dynamic, customizable, and auditable visual explanations, aligning better with scientific practices of modeling and parameter exploration, and means LMMs can create tools for visualization, not just the visualizations themselves.

Furthermore, frameworks like Q-SIT are training LMMs for image quality scoring and interpretation, teaching them low-level visual interpretation skills.[39] While initially focused on general image quality assessment, the principles of teaching models to understand visual attributes have foundational implications for interpreting scientific images, which often contain nuanced details and specific conventions.

The SciVerse benchmark provides critical insights into LMMs' abilities to comprehend multimodal scientific problems, particularly those involving diagrams.[6] Findings from SciVerse indicate that while closed-source LMMs often outperform their open-source counterparts in knowledge comprehension and visual perception in scientific domains, all LMMs currently face challenges with Optical Character Recognition (OCR) and interpreting information that is solely embedded within diagrams.[6] This highlights a crucial area for development if LMMs are to reliably assist with diagram-heavy STEM content.

3.2 LMMs in Making Abstract STEM Concepts Tangible

Many STEM concepts, such as molecular interactions, quantum phenomena, or complex ecological systems, are abstract and challenging for students to grasp. LMMs hold the potential to make these concepts more tangible by generating animations, interactive diagrams, and visual metaphors.[43] For example, an LMM could generate a 3D animation of protein folding or simulate the effects of changing variables in an ecosystem. Multimodal learning analytics (MMLA) principles suggest that integrating diverse data sources (visual, auditory, textual) leads to a richer understanding of learning processes[43]; LMMs can embody this principle by creating content that inherently combines these modalities.

Student engagement can also be enhanced by the aesthetic qualities of AI-generated visuals. Preservice chemistry teachers, for instance, noted that visuals created by GenAI tools were often "attractive, artistic, and interesting".[46] While aesthetic appeal is not a substitute for scientific accuracy, it can serve as an important hook to draw students into complex topics, making initial engagement more likely.

3.3 Challenges: Ensuring Scientific Accuracy and Mitigating Cognitive Biases in AI-Generated Visuals

Despite their potential, LMM-generated visuals come with significant challenges, primarily concerning scientific accuracy and the introduction or amplification of cognitive biases.

The most critical issue is ensuring the scientific accuracy of the visuals. LMMs are known to "hallucinate" or produce plausible-sounding but factually incorrect information, and this extends to visual outputs.[8] Diagrams might contain incorrect relationships, labels, or representations of processes. The study involving preservice chemistry teachers found that while AI-generated visuals were often aesthetically pleasing, they frequently suffered from scientific inaccuracies, pedagogical inappropriateness, or representational limitations (e.g., missing levels of representation like particulate views in chemistry).[46] This dissatisfaction underscores a critical gap: the current inability of many LMMs to consistently produce scientifically sound visuals.

Cognitive biases can also be introduced or perpetuated by AI-generated visuals.[49] LMMs learn from vast datasets, which may contain inherent societal biases related to gender, race, or culture. If these biases are reflected in the visuals—for example, by consistently depicting scientists as male or underrepresenting certain demographic groups in medical illustrations—they can reinforce harmful stereotypes and negatively impact students' perceptions of who belongs in STEM or how scientific concepts apply to diverse populations.[49] Beyond demographic biases, AI visuals might oversimplify complex phenomena or present information in a way that subtly steers interpretation towards a particular (potentially flawed) viewpoint, impacting the development of nuanced scientific understanding.

Another concern is the potential for "model collapse," where LMMs trained extensively on AI-generated content (including visuals) may begin to degrade in quality and accuracy over time, as errors and artifacts in one generation of AI content are learned and amplified by the next.[51] This could lead to a future where the pool of reliable visual information is contaminated.

Finally, the "cognitive ease at a cost" dilemma, identified by Zhai et al.[18] and discussed in [18], is particularly relevant for visuals. If LMM-generated visuals are highly polished and seemingly authoritative, students might accept them uncritically, reducing their cognitive effort but also compromising deeper processing, critical evaluation, and genuine understanding.

3.4 Instructional Design Principles for Effective AI-Generated Visuals

To harness the benefits of LMM-generated visuals while mitigating their risks, careful instructional design is paramount. Applying Mayer's Multimedia Learning Principles[20] to the design and selection of LMM-generated content can help ensure clarity, reduce extraneous cognitive load, and foster generative processing. For example, ensuring that an AI-generated diagram and its accompanying textual explanation are presented contiguously and coherently is vital.

The ARCHED framework's emphasis on human educators as the primary decision-makers is crucial.[27] Educators must retain control over selecting, modifying, and validating AI-generated visual content before presenting it to students, ensuring pedagogical and scientific appropriateness.

Developing students' critical evaluation skills for AI-generated visuals is essential. This includes teaching them to cross-reference information with credible scientific sources, identify potential inaccuracies or biases, and question the representations presented.[51] The dissatisfaction of preservice teachers with AI visuals[46] actually points to an important pedagogical opportunity: LMMs can become powerful tools for developing both teachers' and students' representational competence and critical evaluation skills if their outputs are used as objects of critique rather than infallible sources of truth.

Indeed, AI's "errors" or imperfections in visual generation can be transformed into pedagogical opportunities.[38] Analyzing a flawed AI-generated diagram can spark rich classroom discussions about scientific accuracy, common misconceptions, and the principles of effective scientific representation. This approach shifts the focus from passive consumption of AI content to active, critical engagement.

4 Transforming STEM Learning through LMM-Powered Virtual Experiments

Virtual laboratories have long been a component of STEM education, offering safe, accessible, and repeatable experimental experiences. The advent of LMMs promises to elevate these virtual environments, transforming them into more dynamic, interactive, and inquiry-driven learning tools.

4.1 The Evolution of Virtual Laboratories in STEM Education

Traditional virtual labs typically provide pre-programmed simulations of scientific experiments. Their benefits include enhanced safety (allowing exploration of hazardous procedures without risk), increased accessibility (overcoming limitations of physical lab equipment or location), and repeatability (enabling students to conduct experiments multiple times to observe patterns or test different variables).[9] Platforms like PhET Interactive Simulations have demonstrated the efficacy of such tools in improving conceptual understanding.[53] However, a common limitation of these pre-programmed virtual labs is that they often feature fixed scenarios and offer limited adaptability to individual student inquiries or novel experimental designs.

4.2 LMMs as Engines for Dynamic and Interactive Simulations

LMMs possess the potential to overcome the static nature of traditional virtual labs by serving as engines for more dynamic, interactive, and responsive simulations. Their ability to process natural language prompts, generate code, and reason about complex systems opens new avenues for creating virtual experimental environments. A key area is the development of **Domain-Specific LMMs for Simulations**. For instance, **ChemDFM-X** is a cross-modal dialogue foundation model specifically designed for chemistry.[14] It can handle diverse chemical data modalities, including molecular graphs (2D structures), 3D molecular conformations, tandem mass spectra (MS2), infrared spectra (IR), and various molecular and reaction images. Its capabilities extend to tasks like predicting molecular properties and completing chemical reactions. The architecture of ChemDFM-X, featuring separate encoders for different modalities and a unified LLM decoder (ChemDFM), allows it to integrate and reason over this diverse chemical information. This demonstrates the potential for LMMs to power sophisticated, domain-specific simulations where students could, for example, input a molecular structure and receive predicted properties or simulate reaction pathways. Furthermore, **LLMPhy** offers a framework for complex physical reasoning by synergizing LMMs with physics engines.[58] In a zero-shot manner, the LMM generates code (e.g., Python scripts for a simulator like PyBullet) to iteratively estimate physical hyperparameters of a system, such as friction coefficients or object damping. It achieves this through an implicit analysis-by-synthesis approach, where the LMM proposes parameters, the simulator runs the experiment, and the LMM refines its estimations based on the outcomes. Once these parameters are inferred, LLMPhy can use them to "imagine" or predict the dynamics of the scene, for example, on the TraySim dataset which involves predicting the stability of objects on a tray after an impact. This showcases LMMs not just running simulations, but actively reasoning about and configuring the underlying physics models of those simulations.

These examples illustrate how LMMs can create more open-ended and adaptive virtual experimental environments compared to traditional, pre-scripted ones.[66] Students might be able to define experimental parameters using natural language, or the LMM could dynamically adjust the simulation based on student actions and queries. The true innovation here lies not just in simulating phenomena, but in the potential for LMMs to co-design experiments with students and to explain the simulation's underlying model (even if imperfectly). This fosters deeper model-based reasoning, as students could interact with the LMM to modify parameters, question assumptions embedded in the simulation's code, or even co-create novel experimental setups via natural language, promoting a more active and authentic inquiry role.

4.3 Fostering Scientific Inquiry, Experimentation, and Problem-Solving Skills

LMM-powered virtual labs are well-suited to support Inquiry-Based Learning (IBL) methodologies.[25] By providing flexible and responsive simulated environments, LMMs can empower students to **formulate hypotheses** by allowing them to pose "what if" questions that the LMM translates into simulation parameters. They can also help students **design experiments**, where students could describe an experimental setup in natural language, and the LMM could help configure the virtual lab accordingly, or critique the proposed design for flaws. In terms of **manipulating variables and collecting data**, LMMs can allow for a wider range of variable manipulation than pre-set simulations, and can assist in organizing and presenting the AI-generated data or simulation outcomes in various formats like tables or graphs.[10] Finally, students can **analyze results and draw conclusions** by discussing simulation outcomes with the LMM, which might offer interpretations (to be critically evaluated) or help identify patterns in the data.

Interaction with LMM-driven simulations can also cultivate computational thinking skills.[70] If students are involved in aspects of prompting the LMM for simulation setup, interpreting AI-generated code that drives the simulation, or understanding how the LMM models a particular phenomenon, they are engaging with core computational concepts such as abstraction, modeling, and algorithmic thinking.

4.4 Pedagogical Strategies for AI-Driven Virtual Labs: From Scaffolding to Learning from Failure

Effective use of LMM-powered virtual labs requires thoughtful pedagogical strategies. A key strategy is **Scaffolding Interactions**. AI-based Interactive Scaffolding (AIIS) can be particularly valuable.[72] LMMs can provide different types of scaffolding during virtual experiments, such as conceptual scaffolding to help students understand underlying principles, metacognitive scaffolding to prompt reflection, procedural scaffolding for guidance, and strategic scaffolding to suggest problem-solving approaches. LMMs can offer hints, ask guiding questions, and provide adaptive feedback tailored to student actions within the virtual lab. Another important strategy is **Learning from "Imperfect" Simulations**. The "black-box" nature of some LMMs, if unaddressed, could hinder genuine scientific understanding. LLMPhy, for example, is described as a "black-box optimization framework".[58] If students do not understand why an LMM makes certain choices in setting up or running a simulation, the experience risks becoming a "magic show" rather than a scientific investigation. Therefore, pedagogical approaches must emphasize "opening the black box" or, at minimum, critically probing its outputs and behaviors. The imperfections or unexpected behaviors of AI-driven simulations can be leveraged as powerful teachable moments.[38] When a simulation behaves unexpectedly or produces an "incorrect" result, it provides an opportunity for students to engage in debugging, critical analysis, and a deeper exploration of the underlying models.[74] This process of identifying and rectifying errors mirrors authentic scientific practice. Finally, the **Teacher's Role as Orchestrator** becomes crucial. In LMM-integrated classrooms, the teacher's role shifts from being a primary content deliverer to that of a facilitator of learning experiences.[67] Teachers guide students in their interactions with LMM-powered virtual labs, pose critical questions, help students make sense of complex simulation outputs, and ensure that the technology is used to deepen understanding rather than as a shortcut. Pedagogical models for GenAI in virtual labs emphasize personalized learning paths and the provision of real-time feedback, with the teacher overseeing and augmenting these AI-driven processes.[66]

LMM-powered virtual labs could uniquely bridge the gap between theoretical understanding and practical experimentation by allowing students to seamlessly move between conceptual explanations and their simulated manifestations. A student could ask, "What happens if I double the concentration of this reactant?" and not only observe the change in a simulated chemical reaction but also receive an LMM-generated explanation of why that change occurred, dynamically linking theoretical principles to observable (simulated) outcomes.

The table below provides concrete examples of LMM applications in visual explanations and virtual experiments:

5 Assessment and Evaluation in LMM-Integrated STEM Learning Environments

The integration of LMMs into STEM education necessitates a significant rethinking of assessment and evaluation practices. Traditional methods may prove inadequate when students have access to powerful AI tools capable of generating sophisticated outputs. New frameworks and approaches are emerging that focus on assessing deeper understanding, critical thinking, the process of human-AI collaboration, and metacognitive skills.

5.1 Assessing Student Understanding and Inquiry Skills with AI-Generated Content and Simulations

A primary challenge is ensuring the authenticity of student work when LMMs can produce plausible text, code, visuals, and even simulate experimental results.[78] Assessment strategies must therefore evolve to measure genuine student learning and inquiry skills rather than merely the ability to prompt an AI.

The FACT (Fundamental, Applied, Conceptual, critical Thinking) Assessment Framework offers a balanced approach.[80] Implemented in an Environmental Data Science course, FACT integrates assessments conducted without AI assistance to build and evaluate fundamental coding skills, alongside AI-assisted assignments and projects where students engage with authentic, complex tasks. This dual approach allows educators to gauge foundational knowledge separately from the ability to leverage AI for advanced applications. The study found that AI tools, when coupled with appropriate guidance, improved student performance and enabled them to tackle more complex, real-world problems.[80] This suggests that assessment should not only permit but strategically incorporate AI use for certain tasks, while reserving others for unaided demonstration of core competencies.

LMMs themselves can be used to generate diverse assessment formats, such as multiple-choice questions or analogies, and can even assist in evaluating student presentations based on clarity, understanding, and organization.[78]

Table 3: Exemplar LMM Applications for Visual Explanations and Virtual Experiments in STEM

Application/ Framework	LMM(s) Used (if specified)	Core Capability	STEM Discipline Example	Key Educational Benefit
Visualizing Thought	GPT-4o	Generating conceptual diagrams (e.g., Matplotlib code) to aid reasoning and planning.	Physics (Blocksworld planning), general problem-solving.	Enhances reasoning accuracy, makes AI's intermediate steps transparent.
Q-SIT	Various LMMs	Image quality scoring and interpretation; learning low-level visual interpretation from instruction-based prompts.	General image analysis, potentially applicable to scientific image interpretation (e.g., microscopy, medical imaging).	Foundation for LMMs to emulate human-like visual assessment, crucial for interpreting scientific visual data.
SciVerse Benchmark	Various LMMs	Assessing LMMs' ability to understand multimodal scientific problems (text, diagrams).	Physics, Chemistry, Biology problems involving diagrams and textual information.	Identifies LMM strengths/weaknesses in scientific knowledge comprehension and visual perception (e.g., OCR in diagrams).
ChemDFM-X	ChemDFM (LLM decoder)	Cross-modal dialogue for chemistry; handles molecular graphs, conformations, spectra, images.	Chemistry (molecular property prediction, reaction completion, spectra interpretation).	Enables domain-specific, multimodal chemical simulations and problem-solving; integrates diverse chemical data types.
LLMPhy	OpenAI o1-mini (example)	Zero-shot physical reasoning; LMM generates code to estimate physical parameters by interacting with a physics engine.	Physics (predicting dynamics of objects on a tray under impact, e.g., TraySim dataset).	Allows LMMs to reason about and configure physics simulations, bridging symbolic AI with physical world models.

However, the critical element is the design of these assessments.

Designing effective rubrics for AI-assisted project-based learning in STEM is crucial.[81] Such rubrics should de-emphasize the polish of the final product and instead focus on criteria such as the **creativity and originality** in the student's approach; the **process and iteration**, including how effectively the student used scaffolding steps and incorporated feedback; the depth of the student's **personal reflection and real-world application**; and evidence of the student's **critical use of AI**, showing they evaluated AI-generated content and demonstrated their own understanding.

Ultimately, assessment in the age of LMMs must shift focus. Since LMMs can often generate the products of learning (essays, code, diagrams), evaluation must increasingly target the processes involved: critical thinking, effective prompt engineering, the ability to critically evaluate AI outputs, ethical reasoning, and metacognitive awareness demonstrated during the human-AI interaction. The "how" and "why" of using the LMM become as, if not more, important than the "what" that is produced.

5.2 Frameworks for Evaluating Human-AI Collaboration in Scientific Tasks

As LMMs become collaborative partners in learning, frameworks are needed to evaluate the quality and effectiveness of this human-AI interaction.

The Human-Centric AI-First (HCAIF) Framework emphasizes attribution and reflection as key assessment components.[33] Students are required to clearly document how and where they used GenAI in their work. This transparency allows educators to evaluate not only the final output but also the student's problem-solving capability, critical thinking, and proficiency in co-creating with AI tools. Furthermore, students maintain journals to analyze and reflect on their learning process, the effectiveness of GenAI, its limitations, and its impact on their work. Within HCAIF, AI is utilized for personalized learning, generating feedback, and supporting competency development, while summative assessment considers both the learning process and outcome quality.

The Comprehensive AI Assessment Framework (CAIAF), an evolution of the AI Assessment Scale (AIAS), aims to ethically integrate AI into educational assessments.[83] It provides clear distinctions based on educational levels and incorporates advanced AI capabilities like real-time interactions and personalized assistance. CAIAF promotes responsible AI use by emphasizing ethical principles and offering adaptable strategies, guiding educators beyond simple restrictions towards fostering innovation and academic integrity.

These frameworks underscore the importance of assessing the process of inquiry and collaboration with AI, rather than solely the end product.[79] This might involve evaluating the quality of student prompts, their ability to iterate on AI suggestions, or their justification for accepting or rejecting AI-generated information.

5.3 The Role of Student Metacognition in Navigating LMM Interactions

Metacognition—the ability to understand and regulate one's own learning processes, including self-awareness, planning, monitoring, and evaluation—is crucial for effective learning, particularly when interacting with AI tools.[84] AI can potentially support the development of these skills by providing structured feedback, prompting reflection, and offering insights into learning patterns.

However, a significant challenge emerges from research on the "metacognitive disconnect".[85] Studies indicate that while using AI can improve task performance, users often overestimate their own contribution and understanding, leading to low metacognitive accuracy.[85] Participants in one study using AI for logical reasoning tasks showed improved performance but also a large overestimation of that performance. Paradoxically, higher self-reported AI literacy sometimes correlated with less accurate self-assessment—participants were more confident yet less precise in their performance evaluations.[85]

This metacognitive disconnect poses a substantial threat to genuine learning. If students believe they understand a concept better simply because an AI helped them produce a high-quality output, they may not engage in the deeper cognitive processing necessary for robust learning, nor recognize their actual learning gaps. This could lead to AI inadvertently masking these gaps.

Therefore, pedagogical strategies must explicitly aim to enhance metacognitive monitoring and critical self-reflection when students use LMMs. This might involve tasks that require students to articulate their reasoning before consulting an AI, to compare their own solutions with AI-generated ones and explain discrepancies, or to reflect on the specific ways AI influenced their thinking and final product. Effective assessment in this context will require a "triangulation" approach: combining student self-reflections (as in HCAIF[33]), AI-supported tasks (as in FACT[80]), and AI-free tasks (also in FACT[80]) to obtain a more holistic and accurate view of student competence and understanding.

6 Navigating Ethical Dilemmas and Practical Challenges

The integration of LMMs into STEM education, while promising, is fraught with ethical dilemmas and practical challenges that demand careful consideration and proactive strategies. These range from ensuring the scientific validity of AI-generated content to addressing bias, maintaining academic integrity, and managing cognitive impacts.

6.1 Upholding Scientific Rigor: Verification of LMM-Generated Explanations and Simulations

A paramount concern in STEM is the accuracy and reliability of information. LMMs, despite their sophistication, can generate explanations, visualizations, and simulation outputs that are inaccurate, based on "hallucinated" data, or subtly biased.[8] They may cite non-existent sources or misinterpret scientific principles.[51] The findings from the SciVerse benchmark, which revealed LMMs' difficulties with accurate visual interpretation and OCR in scientific diagrams[6], underscore this challenge. Similarly, preservice chemistry teachers found AI-generated visuals often lacked scientific rigor despite their aesthetic appeal.[46]

This necessitates a strong emphasis on verification. Students and educators must be equipped with the skills and tools to critically evaluate and fact-check AI-generated scientific content, cross-referencing it with established, credible sources. This challenge of verification, however, can be reframed as a pedagogical opportunity. The process of scrutinizing AI outputs can itself teach critical thinking, information literacy, and the nature of scientific evidence in the digital age. When students are tasked with identifying flaws in an AI-generated explanation or simulation, they are forced to engage more deeply with the underlying scientific concepts and the criteria for trustworthy claims.

6.2 Addressing Algorithmic Bias, Equity, and Accessibility in LMM Deployment

LMMs learn from vast datasets, which often reflect existing societal biases related to gender, race, culture, and other demographic factors.[86] If these biases are embedded in the LMMs, they can be perpetuated or even amplified in educational content. In STEM, this could manifest as stereotypical representations of scientists, biased interpretations of data related to certain populations, or inequitable performance of AI tools for different student groups.[49] Addressing algorithmic bias is crucial not only for fairness but also for scientific validity, as biased LMMs could misrepresent scientific phenomena or steer inquiry in skewed directions.

Equity issues also extend to access. The most powerful LMMs are often proprietary and may require subscriptions, creating a divide between students who can afford them and those who rely on free, potentially less capable, versions.[86] Differences in digital literacy can further exacerbate these inequities. Educational institutions must strive to provide equitable access to LMM tools and the training needed to use them effectively, ensuring that LMM integration does not widen existing educational disparities.

6.3 Maintaining Academic Integrity with Advanced LMM Capabilities

The ability of LMMs to generate human-like text, code, and other outputs raises significant concerns about academic integrity.[86] The risk of students submitting AI-generated work as their own, or using AI to complete assignments without genuine understanding, is a major challenge for educators.

Addressing this requires a multi-pronged approach. Clear institutional policies regarding the permissible and ethical use of AI in academic work are essential, emphasizing transparent disclosure of AI assistance and the importance of original thought.[86] Assessment methods may need to be redesigned to be "AI-resistant," focusing on tasks that require higher-order thinking, in-class performance, or process-oriented evaluation rather than solely on final products that AI can easily generate.[88] As discussed previously, shifting the focus of assessment to the student's process of interaction with AI, their critical evaluation of its outputs, and their ability to synthesize information from multiple sources (including AI) can help maintain academic integrity.

6.4 Managing Cognitive Load and Preventing Over-Reliance

While LMMs can assist in complex tasks and potentially reduce certain aspects of cognitive load, there is a significant risk of "cognitive offloading," where students become overly dependent on AI tools to do the thinking for them.[86] This over-reliance can hinder the development of essential cognitive skills, critical thinking, and deep conceptual understanding. The "cognitive ease at a cost" phenomenon, where LLMs make tasks feel easier but lead to shallower processing and poorer reasoning[18], highlights this danger. If students consistently use LMMs as a crutch rather than a tool for augmentation, their own problem-solving and analytical abilities may atrophy.

Pedagogical strategies must be designed to encourage active engagement and critical thinking even when LMMs are used. This involves structuring tasks so that AI assists with certain components (e.g., data processing, initial drafting) while requiring students to perform the core intellectual work (e.g., hypothesis formulation, critical analysis, synthesis of ideas). The tension between leveraging LMMs for efficiency and accessibility versus ensuring deep learning and skill development represents a central ethical and pedagogical balancing act for educators. The goal is to find a "sweet spot" where LMMs scaffold and augment human cognition and effort, rather than supplanting them.

7 Future Trajectories: Research and Practice for LMMs in STEM Education

The rapid advancement of LMMs necessitates a forward-looking perspective on their role in STEM education. Future developments will likely focus on empowering educators, designing more sophisticated AI-driven learning environments, reimagining the scientific method through human-AI collaboration, and addressing grand challenges in research.

7.1 Empowering Educators: Professional Development and Classroom Orchestration Strategies

For LMMs to be effectively integrated into STEM classrooms, educators must be equipped with the necessary knowledge and skills. This requires robust professional development (PD) programs that go beyond basic tool usage, focusing on LMM capabilities, inherent limitations, pedagogical implications, and ethical considerations.[67] The Shark AI project, which provides PD for middle school teachers on using AI in paleontology, serves as an example of how teachers can be supported in implementing AI-infused curricula.[93] Such PD should emphasize peer-to-peer learning and address common misconceptions about AI.

Table 4: Ethical Challenges and Proposed Mitigation Strategies for LMMs in STEM Education

Ethical Challenge	**Description of Challenge in STEM Context**	**Proposed Mitigation Strategies (Pedagogical, Technical, Policy)**
Scientific Inaccuracy/ Hallucinations	LMMs generating factually incorrect scientific explanations, flawed diagrams, or unrealistic simulation outputs; citing non-existent sources.	Pedagogical: Teach critical evaluation, source verification, cross-referencing. Use errors as learning opportunities. Technical: Develop domain-specific LMMs with curated STEM knowledge. Implement better fact-checking mechanisms within LMMs. Policy: Guidelines for validating AI-generated content.
Algorithmic Bias	LMMs perpetuating gender, racial, or cultural biases in STEM representations (e.g., images of scientists, medical data interpretation), leading to skewed understanding or inequity.	Pedagogical: Educate on AI bias, encourage critical analysis of AI outputs for bias. Technical: Use diverse and representative training data. Implement bias detection and mitigation algorithms. Policy: Promote development of fair and unbiased AI tools for education.
Academic Misconduct	Students submitting AI-generated work as their own; plagiarism; lack of genuine understanding despite polished outputs.	Pedagogical: Focus on process over product in assessments. Teach ethical AI use and proper attribution. Design AI-resistant assignments. Technical: Improve AI-detection tools (though with limitations). Policy: Clear institutional policies on AI use, academic integrity, and consequences of misuse.
Cognitive Over-reliance/ Deskilling	Students becoming overly dependent on LMMs, leading to reduced critical thinking, problem-solving skills, and shallower learning ("cognitive offloading").	Pedagogical: Design tasks requiring active student engagement and critical thinking beyond AI capabilities. Teach metacognitive strategies for AI use. Balance AI-assisted and AI-free tasks. Policy: Emphasize development of core cognitive skills alongside AI literacy.
Access and Equity	Disparities in access to powerful LMMs (paid vs. free), necessary hardware, internet connectivity, and digital literacy, potentially widening educational gaps.	Pedagogical: Provide training for all students on available tools. Technical: Promote open-source LMM development. Design tools requiring less computational power. Policy: Ensure equitable access to technology and AI tools in educational institutions. Fund digital literacy programs.
Data Privacy and Security	Collection and use of student data by LMMs, raising concerns about privacy, surveillance, and potential misuse of sensitive information.	Technical: Use LMMs with strong data protection features. Anonymize data where possible. Policy: Adherence to data privacy regulations (e.g., GDPR, FERPA). Transparent institutional policies on data handling by AI tools. Obtain informed consent.

The role of the teacher is evolving into that of a "classroom orchestrator".[76] Instead of being the sole source of information, teachers will facilitate and guide student interactions with LMMs, help students critically interpret AI-generated outputs, and ensure that technology serves pedagogical goals. This involves strategies like using AI as a debate partner, an assistant for mock interviews, or a personalized study buddy, always with an emphasis on human review, bias checking, and contextualization of AI-generated content.[76]

7.2 Designing Next-Generation AI-Driven Learning Environments for Deeper Scientific Understanding

Future learning environments will likely see LMMs embedded more deeply into core curricula, aiming to build AI fluency for all students, regardless of their major.[95] This involves moving beyond standalone AI tools to integrated platforms that support diverse learning activities. Emerging tools like Curipod (for interactive lessons), MagicSchool.ai (for teacher assistance), Labster (for virtual labs), Tynker with AI (for coding), and Exploratorium AI Labs exemplify this trend towards more comprehensive AI-driven learning environments.[38]

A key direction in designing these environments is to prioritize open-ended, creative problem-solving and student agency, rather than relying on rigid, AI-tutoring systems that dictate the learning path.[96] LMMs should function more like collaborative peers or sophisticated reference materials that students can use on their own terms to explore, create, and discover. The most effective future LMM integration will likely involve domain-specific LMMs that are fine-tuned with pedagogical knowledge, rather than relying solely on general-purpose LMMs for specialized STEM education tasks. General LMMs often lack the conceptual depth and accuracy required for specific STEM fields.[97] Domain-specific models, such as ChemDFM-X for chemistry[56], show greater promise. Integrating established pedagogical frameworks directly into these models would represent a significant advancement, ensuring that the AI not only understands the content but also how to present and interact with it in a pedagogically effective manner.

7.3 The Evolving Scientific Method: Long-Term Vision for Human-AI Partnership in Discovery and Learning

LMMs have the potential to transform not only how science is learned but also how it is done. The long-term vision involves AI as a collaborative partner, augmenting human capabilities in scientific discovery, data analysis, hypothesis generation, and even the articulation of scientific findings.[96] LMMs could assist researchers and students in identifying patterns in large datasets, suggesting novel research questions, or drafting initial sections of scientific papers.[100]

This human-AI partnership in science implies a fundamental shift in scientific training. Students will need to learn not just the principles of science, but also how to conduct scientific inquiry collaboratively with AI. This requires new curricula focused on effective human-AI scientific workflows, including skills in advanced prompt engineering, critical evaluation of AI-generated hypotheses or analyses, and ethical considerations of AI in research. For such collaboration to be fruitful, there is a need for shared understanding and contextual awareness between humans and AI agents, where AI can interpret human inputs effectively and humans can understand and act upon AI outputs.[100]

7.4 Grand Challenges and Priority Research Areas

Several grand challenges and priority research areas must be addressed to realize the full potential of LMMs in STEM education. A major challenge is ensuring the accuracy and depth of generated content, as general-purpose LMMs often lack conceptual rigor, necessitating the development of domain-specific models with integrated pedagogical frameworks.[97] Further research is needed to understand the impact on higher-order thinking skills to ensure LMMs augment, rather than diminish, critical analysis and creativity. The assessment validity of AI-assisted methods requires significant investigation to accurately measure student learning.[102] Overcoming issues of equity and access to powerful LMMs and the necessary hardware remains a significant hurdle.[97] The epistemology of AI-driven science is another critical area, as students need to understand the nature of AI-generated knowledge.[32] Finally, a significant grand challenge lies in explainability and transparency, as developing LMMs that can explain their reasoning and articulate uncertainties is crucial for fostering trust and enabling critical evaluation by learners, moving beyond problematic "black box" models.[107]

Addressing these challenges will require interdisciplinary collaboration among AI researchers, STEM educators, learning scientists, and policymakers.

8 Conclusion

8.1 Recapitulation of LMMs' Transformative Potential in STEM Education

Large Multimodal Models stand at the cusp of significantly reshaping STEM education. Their capacity to process and generate information across diverse modalities—text, image, code, and simulation data—offers unprecedented opportunities. This review has highlighted how LMMs can enhance visual explanations by creating dynamic, interactive, and contextually rich scientific visualizations, thereby making abstract concepts more accessible. Furthermore, LMMs show considerable promise in powering sophisticated virtual experiments, allowing for more open-ended, inquiry-driven, and personalized laboratory experiences that can foster critical scientific skills. From generating conceptual diagrams that elucidate reasoning processes to enabling complex simulations in fields like chemistry and physics, LMMs offer tools that can deepen student engagement and understanding in STEM.

8.2 Key Implications for Educational Stakeholders

The integration of LMMs carries profound implications for all stakeholders in the educational ecosystem. For educators, there is an urgent need to develop new pedagogical skills, adapt teaching methodologies, and cultivate a keen awareness of the ethical dimensions of AI use, shifting their role towards that of a facilitator and orchestrator.

For students, LMMs present opportunities for more engaging and personalized learning but also pose risks of over-reliance and the imperative to develop robust critical AI literacy. For **developers**, the onus is on creating LMM tools that are scientifically accurate, free from harmful biases, and designed with sound pedagogical principles and transparency. Finally, for policymakers, clear guidelines for the ethical and effective use of LMMs are needed, along with investment in research, teacher professional development, and initiatives to ensure equitable access.

Table 5: Leading Journals for Research on AI in STEM Education

Journal Name	**Scope Relevant to AI/STEM/EdTech**	**Metric (h5-index, etc.)**
Computers & Education	Digital technology to enhance education, pedagogical uses of digital technology, learning and teaching implications. High encouragement for empirical research on diversity, equity, and inclusion.	154
Education and Information Technologies	Educational technologies, information technologies in education.	112
British Journal of Educational Technology (BJET)	Theory, applications, and development of learning technology, ICT in education and training.	101
Interactive Learning Environments	Design and use of interactive learning environments, human-computer interaction in education.	85
Computer Assisted Language Learning	AI in language learning, technology-enhanced language teaching. (Relevant for AI pedagogy aspects).	79
International Journal of Educational Technology in Higher Education	Technology in higher education, innovative uses of EdTech.	77
Educational Technology Research and Development (ETR&D)	Research and development in educational technology, instructional design, learning sciences.	72
Journal of Computer Assisted Learning (JCAL)	Computer-assisted learning, collaborative learning, AI in education, open and networked learning.	63
Computers and Education: Artificial Intelligence	AI applications in education, intelligent tutoring systems, learning analytics, ethical issues of AI in education.	57
Journal of Writing Research (JoWR)	Cognitive/social processes of writing, learning/teaching writing, technology in writing. (Relevant for AI in scientific communication).	Top 10% Education (Scopus)
Computers and Composition: An International Journal	Computers in writing classes, programs, research; digital writing, electronic literacy, multimedia composition, technology in writing instruction. (Relevant for multimodal aspects and scientific writing with AI).	N/A (Impactful in its field)
Educational Technology & Society	Educational technology, active/experiential/cooperative learning, teaching methods, multimedia, AI in education. Open access.	2.633 (2021 Impact Factor)

The successful integration of LMMs into STEM education is not a purely technological endeavor but a complex socio-technical challenge. It demands deliberate, ethical, and pedagogically informed choices from all involved parties. Passive adoption of these powerful tools is insufficient; active, critical engagement is required to navigate their potential and pitfalls.

8.3 A Call for Responsible Innovation and Collaborative Research

The journey of integrating LMMs into STEM education is just beginning. While the potential is immense, so are the unknowns and the responsibilities. There is a pressing need for sustained, rigorous research into the efficacy, equity, and ethical implications of these technologies in diverse STEM learning contexts. This research must be interdisciplinary, bringing together AI developers, STEM experts, learning scientists, ethicists, and classroom educators.

Collaboration is key to ensuring that LMMs are developed and deployed in ways that genuinely augment human intellect, foster deep conceptual understanding, and prepare students for a future where AI is an integral part of scientific inquiry and professional practice. The ultimate measure of LMMs' success in STEM education will not be their technical sophistication alone, but their ability to cultivate critical thinking, creativity, problem-solving skills, and a robust epistemic understanding of science in an increasingly AI-mediated world. These higher-order goals must remain central as we navigate the transformative landscape of LMMs in education.

To Cite This Article Changkui LI. (2025). Applications of Large Multimodal Models (LMMs) in STEM Education: From Visual Explanations to Virtual Experiments. *Artificial Intelligence Education Studies*, *1*(2), 1-18. https://doi.org/10.6914/aiese.010201

Reference

[1] Wu, S., Yin, S., & Li, L. (2024). On opportunities and challenges of large multimodal foundation models. *Fundamental Research, 4*(2), 241 244.

[2] Cem, B. (2024, May 15). Large Multimodal Models (LMMs) vs LLMs in 2025. *AIMultiple*.

[3] STEM Education Center. (n.d.). Navigating the Future: The Role of AI in Transforming K-12 Science Education Globally. *stemedcenter.org*.

[4] Google Cloud. (n.d.). Multimodal AI. *Google Cloud*.

[5] Zhy, T. (2024, April 19). Multimodal AI: A guide to open-source Vision Language Models. *BentoML*.

[6] Wang, T., Zhang, N., Zhang, Z., Zhao, Y., Liu, Z., Wang, Z., & Liu, Y. (2025). SciVerse: Unveiling the Knowledge Comprehension and Visual Reasoning of LMMs on Multi-modal Scientific Problems. *arXiv*.

[7] Di Cecco, D. (2024, August 28). Multimodal Large Language Models. *Neptune.ai*.

[8] Yendi, H. M., & Susilowati, E. (2024). Integrating generative AI into STEM education: Insights from science and mathematics teachers. *International Electronic Journal of Mathematics Education, 19*(3), em0779.

[9] Oregon State University. (2024, June 12). Engaging online labs and courses grounded in pedagogy launch STEM students to success. *College of Science*.

[10] Tretinjak, M. F. (n.d.). Hands-on-Learning in Teaching of Science: Virtual Labs and Simulations. *Innovative Pedagogies in Science Education, Pressbooks*.

[11] Labster. (n.d.). Immersive STEM Learning Experiences. *Labster*.

[12] Stemble. (n.d.). Interactive Virtual Labs. *Stemble Learning*.

[13] Yang, Z., Zhang, J., Li, L., Zhang, A., Zhang, A., & Chen, H. (2025). Visualizing Thought: Conceptual Diagrams Enable Robust Planning in LMMs. *arXiv*.

[14] Li, Z., Guo, Z., Huang, Z., Zhu, R., & Liu, Z. (2024). ChemDFM-X: Towards Large Multimodal Model for Chemistry. *arXiv*.

[15] Cowling, M., & Birt, J. (2024). Shaping integrity: why generative artificial intelligence does not have to undermine education. *Frontiers in Artificial Intelligence,* 7.

[16] Koanly. (2024, August 22). The power of Generative Artificial Intelligence in the Constructivist Model in education. *Koanly*.

[17] Newton, P. M. (2024). Challenging Cognitive Load Theory: The Role of Educational Neuroscience and Artificial Intelligence in Redefining Learning Efficacy. *Education Sciences, 14*(4), 361.

[18] Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., ... & Amazeen, M. A. (2022). Cognitive ease at a cost: LLMs reduce mental effort but compromise performance in a headline verification task. *Opus, University of Augsburg*.

[19] Werder, K., Ramesh, V., Zhang, L., & Wang, J. (2024). Cognitive Load Dynamics in Generative AI-Assistance: A NeuroIS Study. *Proceedings of the International Conference on Information Systems (ICIS)*.

[20] Digital Learning Institute. (2023, April 25). Mayer' s 12 Principles of Multimedia Learning. *Digital Learning Institute*.

[21] Gordon, C. (2024, March 6). Supporting learning with AI-generated images: A research-backed guide. *MIT Sloan Education Technology*.

[22] Illumina Interactive. (2023, October 19). Revisiting Multimedia Learning Theory: Fostering Generative Processing in eLearning. *Illumina Interactive*.

[23] Eduaide.Ai. (n.d.). Richard Mayer on Multimedia Learning, Transfer, and the Future of Educational Psychology (Ep. 001). *Eduaide.Ai*.

[24] Conrad, S., La, D., & Laurillard, D. (2025, April 21). Mapping a Multidimensional Framework for GenAI in Education. *EDUCAUSE Review*.

[25] Turnitin. (n.d.). How inquiry-based learning develops critical thinking in STEM subjects. *Turnitin*.

[26] Lee, S., & Kim, M. (2024). Inquiry-based Artificial Intelligence Curriculum for Upper Elementary Students: A Design Case of PrimaryAI. *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[27] Afzaal, M., Bou-Harb, E., & Ghani, S. (2025). ARCHED: A Human-Centered Framework for Transparent, Responsible, and Collaborative AI-Assisted Instructional Design. *arXiv*.

[28] Warschauer, M., & O' Byrne, W. I. (2024). AI-writing tools in education: if you can't beat them, join them. *Journal of Learning and Teaching in Digital Age, 6*(1), 1 11.

[29] Ziegler, N. (2024). Artificial intelligence for language learning: Entering a new era. *Language Learning & Technology, 28*(2), 1 9.

[30] ResearchGate. (n.d.). Waverly Tseng's research works. *ResearchGate*.

[31] Bowdoin College Library. (n.d.). Readings and Resources - Artificial Intelligence at Bowdoin. *Bowdoin College*.

[32] Dai, Y., & Li, T. (2024). Unpacking Epistemic Insights of Artificial Intelligence (AI) in Science Education: A Systematic Review. *Journal of Science Education and Technology, 33*(4), 481-496.

[33] AACSB International. (2025, February 11). A Framework for Human-Centric AI-First Teaching. *AACSB Insights*.

[34] Georgetown University Writing Program. (n.d.). Principles of Critical AI Literacy. *Georgetown University*.

[35] Georgetown University Writing Program. (n.d.). Principles of Critical AI Literacy. *Georgetown University*.

[36] Notre Dame Learning. (2024, April 3). Building Critical AI Literacy in the Humanities Classroom. *University of Notre Dame*.

[37] Georgetown University Writing Program. (n.d.). Guidelines for Teaching. *Georgetown University*.

[38] Force, C. (2024, May 22). How AI Is Transforming Critical Thinking in Science Education. *NSTA Blog*.

[39] Wu, Y. (2025, May 27). Teaching LMMs for Image Quality Scoring and Interpreting. *SlideShare*.

[40] Yang, Z., Zhang, J., Li, L., Zhang, A., Zhang, A., & Chen, H. (2025). Visualizing Thought: Conceptual Diagrams Enable Robust Planning in LMMs. *arXiv*.

[41] Yang, Z., Zhang, J., Li, L., Zhang, A., Zhang, A., & Chen, H. (2025). Visualizing Thought: Conceptual Diagrams Enable Robust Planning in LMMs. *arXiv*.

[42] Wang, T., Zhang, N., Zhang, Z., Zhao, Y., Liu, Z., Wang, Z., & Liu, Y. (2025). SciVerse: Unveiling the Knowledge Comprehension and Visual Reasoning of LMMs on Multi-modal Scientific Problems. *arXiv*.

[43] Liu, J., Li, H., Liu, Y., & Zhang, J. (2025). A Comprehensive Review of Multimodal Analysis in Education. *Applied Sciences, 15*(11), 5896.

[44] Edutopia. (2024, January 26). The Power of Multimodal Learning (in 5 Charts). *Edutopia*.

[45] NSTA. (n.d.). Teaching With Simulations. *NSTA*.

[46] Tang, S. Y., & Tang, H. (2024). Generative Artificial Intelligence (GenAI) as the Artist of Chemistry and the Mentor for Chemistry Education. *Journal of Chemical Education*.

[47] Institute of Education Sciences. (n.d.). Scientific Misconceptions: From Cognitive Underpinning to Educational Treatment. *IES*.

[48] Streeter, I., & Ravan, M. (2021). Examining data visualization pitfalls in scientific publications. *PLOS ONE, 16*(10), e0258832.

[49] Nicastro, M., & Charytan, D. M. (2025). From data to perception: visualizing bias in artificial intelligence-generated images. *European Heart Journal, 46*(19), 1781 1783.

[50] Sena, J., van der Goot, A., & Bluemke, D. A. (2024). Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. *European Radiology*.

[51] Columbia College (BC). (n.d.). Assessing AI-Generated Content. *LibGuides*.

[52] NorthWest Arkansas Community College Library. (n.d.). Ethical AI: Using AI for Your Research: Verify AI-Generated Data. *NWACC Library*.

[53] Perkins, K., Adams, W., Dubson, M., Finkelstein, N., Reid, S., Wieman, C., & LeMaster, R. (2006). PhET: Interactive Simulations for Teaching and Learning Physics. *The Physics Teacher, 44*(1), 18-23.

[54] PhysPort. (n.d.). Methods and Materials: PhET Interactive Simulations. *PhysPort*.

[55] Liu, S. (2025). Exploration of the Path of Artificial Intelligence Enabling Science Education. *Open Journal of Social Sciences, 13*, 463-470.

[56] Li, Z., Guo, Z., Huang, Z., Zhu, R., & Liu, Z. (2024). ChemDFM-X: towards large multimodal model for chemistry. *Science China Information Sciences, 67*(11), 220109.
[57] Merchant, A., et al. (2025). An Artificial Intelligence Research Roadmap for Molecules, Materials and Proteins. *ChemRxiv*.
[58] Corcodel, R., Cherian, A., & Freeman, D. (2024). LLMPhy: Complex Physical Reasoning Using Large Language Models and World Models. *arXiv*.
[59] Naimi, A., & Naimi, E. (2024). Large Physics Models: Towards a collaborative approach with Large Language Models and Foundation Models. *ResearchGate*.
[60] Corcodel, R., Cherian, A., & Freeman, D. (2024). LLMPhy: Complex Physical Reasoning Using Large Language Models and World Models. *arXiv*.
[61] Google Scholar. (n.d.). Radu Corcodel. *Google Scholar*.
[62] Corcodel, R., Cherian, A., & Freeman, D. (2024). LLMPhy: Complex Physical Reasoning Using Large Language Models and World Models. *OpenReview*.
[63] AZoAI. (2024, November 18). LLMPhy Revolutionizes Physical Reasoning by Combining AI and Physics Simulation. *AZoAI*.
[64] Papers With Code. (n.d.). Anoop Cherian. *Papers With Code*.
[65] Google Scholar. (n.d.). Educational Technology - Google Scholar Metrics. *Google Scholar*.
[66] Qatar University. (2024, May 14). Exploring Generative AI in Education with Innovative Applications and Transformative. *Qatar University*.
[67] EDC & CADRE. (2025, January). Generative AI in STEM Teaching: Opportunities and Tradeoffs. *CADRE*.
[68] Jeong, H., & Hmelo-Silver, C. E. (2023). Artificial Intelligence-Based Scientific Inquiry. *ResearchGate*.
[69] Institute of Education Sciences. (n.d.). Investigating the Use of Virtual Labs to Promote Chemical Reasoning and Practice Skills. *IES*.
[70] Kong, S.-C., & Sheldon, J. (2021). *Computational Thinking Education in K-12: Artificial Intelligence Literacy and Physical Computing*. MIT Press.
[71] Raspberry Pi Foundation. (2024, May 29). Teaching about AI in K 12 education: Thoughts from the USA. *Raspberry Pi Foundation*.
[72] Wu, Y. (2025). The effects of artificial intelligence-based interactive scaffolding on EFL learners' academic writing performance, writing anxiety, and cognitive load. *Interactive Learning Environments*.
[73] Wang, R., & Tan, T. (2025). Scaffolding Creativity: Integrating Generative AI Tools and Real-world Experiences in Business Education. *arXiv*.
[74] Venkatraman, S. (2025). AI-Powered Debugging: Exploring Machine Learning Techniques for Identifying and Resolving Software Errors. *PhilArchive*.
[75] Focalx. (n.d.). AI Debugging: Identifying and Fixing Model Errors. *Focalx*.
[76] Getting Smart. (2025, January 7). How Teachers Can Orchestrate a Classroom Filled with AI Tools. *Getting Smart*.
[77] Jurisz, S. (2024, February 21). 5 Engaging AI Classroom Activities to Try With Your Students. *Edutopia*.
[78] Hurix Digital. (2024, August 28). Create Smarter Assessments Using AI to Improve Student Outcomes. *Hurix Digital*.
[79] UNSW Sydney. (n.d.). Solving AI challenges in teaching. *UNSW Sydney*.
[80] Le, V. A., Puc, M., & Zupanc, K. (2025). Balancing AI-assisted learning and traditional assessment: The FACT assessment in environmental data science education. *Frontiers in Education, 10*.
[81] CUNY. (n.d.). Build Project-Type Assessments with AI Guidance. *Hostos Community College EdTech*.
[82] IDEEAS Lab. (n.d.). Designing Rubrics for AI-Enabled Assignments. *IDEEAS Lab*.
[83] Adigun, J. O. (2025). Comprehensive AI assessment framework: Enhancing educational evaluation with ethical AI integration. *IAU International Journal of Educational Technology, 3*(1), 1-15.
[84] Hernández-de-Menéndez, M., & Morales-Menendez, R. (2025). Building Metacognitive Skills Using AI Tools to Help Higher Education Students Reflect on Their Learning Process. *IEEE RITA, 18*(1), 16-23.
[85] Liu, Y., Lu, Y., Niu, Y., & Li, G. (2024). Large Language Models for Education: A Survey and Outlook. *arXiv*.
[86] Demetriou, N., & Vardi, M. Y. (2025). Large Language Models in Computer Science Classrooms: Ethical Considerations. *Applied Sciences, 15*(4), 1793.
[87] Litslink. (2024, May 7). Generative AI in Education: The Impact, Ethical Considerations, and Use Cases. *Litslink*.

[88] Chan, C. K. Y. (2025). Redesigning Assessments for AI-Enhanced Learning: A Framework for Educators in the Generative AI Era. *Education Sciences, 15*(2), 174.

[89] EDUCAUSE. (2023). Examining the Impact of Large Language Models on Higher Education: Ethical Considerations. *EDUCAUSE*.

[90] ResearchGate. (n.d.). How can generative AI tools be effectively integrated into higher education to enhance critical thinking, academic integrity, rather than hinder them? *ResearchGate*.

[91] Smith, A., & Jones, B. (2025). The Impact of Artificial Intelligence (AI) on Students' Academic Development. *Education Sciences, 15*(3), 343.

[92] Toddle. (n.d.). The Ultimate Guide to Evaluating AI Tools in Your School. *Toddle Learn*.

[93] National Science Foundation. (2024, March 14). Preparing science educators to use and teach AI in the classroom. *NSF*.

[94] New York Academy of Sciences. (2024, February 12). Bringing Science to Life with Artificial Intelligence. *NYAS*.

[95] The Ohio State University. (2024, May 15). Ohio State launches bold AI Fluency initiative to redefine learning and innovation. *The Ohio State University*.

[96] Northwestern University. (2025, May 7). Exploring AI in Education Through the Lens of the Learning Sciences. *Northwestern University*.

[97] CADRE. (n.d.). Artificial Intelligence in STEM Education Research. *CADRE*.

[98] Wang, Z., Ber Ziv, I., & Gill, B. (2025). Human-AI Collaboration or Academic Misconduct? Measuring AI Use in Student Writing Through Stylometric Evidence. *arXiv*.

[99] Zhang, J., et al. (2024). XtraGPT: LLMs for Human-AI Collaboration on Controllable Academic Paper Revision. *arXiv*.

[100] Mata, C. P., et al. (2025). A vision of human AI collaboration for enhanced biological collection curation and research. *BioScience*.

[101] Lawrence Livermore National Laboratory. (2025, April 24). Stanford professor discusses future of visually intelligent machines and human-AI collaboration. *LLNL*.

[102] The 74. (2024, April 17). STEM' s Grand Challenges —And Opportunities. *The 74*.

[103] Jumper, J., & Hassabis, D. (2022). The Epistemology of AI-driven Science: The Case of AlphaFold. *PhilSci-Archive*.

[104] Adle, A. (2025). AI and Epistemic Agency: How AI Influences Belief Revision and Its Normative Implications. *Social Epistemology*.

[105] Kelly, K. T. (2014). Computational Epistemology. In *Mainstream and Formal Epistemology* (pp. 165–188). Cambridge University Press.

[106] Grote, G. (2019). Methodology and Epistemology of Computer Simulations and Implications for Science Education. *ResearchGate*.

[107] Institute of Education Sciences. (n.d.). National Center on Generative AI for Uplifting STEM+C Education (GENIUS Center). *IES*.

[108] Google Scholar. (n.d.). Educational Technology - Google Scholar Metrics. *Google Scholar*.

[109] Anglin, G. J. (n.d.). The Journals in the field of Instructional Technology. *designer.50g.com*.

[110] Elsevier. (n.d.). Journals in Education. *Elsevier*.

[111] Google Scholar. (n.d.). Educational Technology - Google Scholar Metrics. *Google Scholar*.

[112] Golonka, E., & Torkleson, A. (Eds.). (2025). *The Routledge Handbook of Technological Advances in Researching Language Learning*. Routledge.

[113] Journal of Writing Research. (n.d.). Home. *Journal of Writing Research*.

[114] Journal of Writing Research. (n.d.). Aims and scope. *Journal of Writing Research*.

[115] S-Logix. (n.d.). Computers and Composition Elsevier Journal Impact Factor. *S-Logix*.

[116] Research.com. (n.d.). Journal of Educational Technology and Society - Impact Factor & Score 2025. *Research.com*.

[117] Wikipedia. (2024, May 24). Educational Technology & Society. *Wikipedia*.

Editor *Sophia LI wtocom@gmail.com*