International Journal Of AI-Enhanced Visual Learning
Vol. 1 No. 1 (2026)

BEYOND THE TEXTBOOK: AI-GENERATED VISUALS AND THE FUTURE OF MULTIMODAL LEARNING

Usama Rasheed
Mphil International Relations, (IR), University of Okara, Okara, Pakistan
Email: chusamarasheed096@gmail.com

Abstract
The rapid development of generative artificial intelligence has changed the ways educational content can be produced, presented, and experienced. Among the most significant developments is the ability of artificial intelligence systems to generate diagrams, illustrations, conceptual images, simulations, infographics, and other forms of visual material from natural-language prompts. This development has particular significance for multimodal learning, in which learners encounter knowledge through combinations of written language, visual representation, audio, movement, and interactive media. This article examines the pedagogical possibilities and limitations of AI-generated visuals in contemporary education. It argues that the educational value of artificial intelligence does not reside simply in its capacity to produce attractive images; rather, its significance depends upon the relationship between visual representation, instructional purpose, cognitive processing, learner participation, and teacher judgment. Drawing on scholarship concerning multimedia learning, multimodality, generative artificial intelligence, and educational technology, the study explores how AI-generated visuals can supplement conventional textbooks, support the explanation of abstract concepts, facilitate contextualized learning, encourage multimodal composition, and enable more adaptable instructional materials. Recent empirical research has reported advantages for multimodal instruction incorporating AI-generated images, although such findings do not necessarily establish that AI-generated images themselves are superior to conventional visual resources. The article also considers challenges involving factual accuracy, cognitive overload, cultural representation, authorship, bias, privacy, accessibility, and teacher oversight. It proposes a human-centred model in which AI functions as a visual production and adaptation assistant rather than an autonomous teacher. The study concludes that the future of multimodal learning should not be understood as a replacement of textbooks by artificial intelligence but as a movement toward more flexible learning environments in which text, image, sound, interaction, and human interpretation work together.

Keywords: *artificial intelligence, AI-generated visuals, multimodal learning, educational technology, visual pedagogy, generative AI, multimedia learning, digital education, instructional design, human-centred AI*

1. Introduction

The textbook has long occupied a central position in formal education. It provides organized knowledge, establishes curricular sequences, supplies explanations, and offers students a relatively stable reference point. Yet contemporary learning increasingly occurs beyond the printed page. Digital classrooms, interactive presentations, educational videos, simulations, online repositories, virtual environments, infographics, and social-media-based educational resources have expanded the range of modes through which knowledge can be communicated. The emergence of generative artificial intelligence adds another significant dimension to this transformation because teachers

and learners can now request the production or adaptation of visual material using ordinary language.

Generative AI can produce or transform content across multiple modes, including text, images, audio, video, and code. UNESCO's guidance on generative AI in education emphasizes that these capabilities have implications not only for classroom activities but also for curriculum design, assessment, research, regulation, and understandings of knowledge itself. The organization nevertheless stresses a human-centred approach that protects human agency, inclusion, equity, cultural and linguistic diversity, and meaningful participation.

The educational importance of AI-generated visuals arises partly from a longstanding observation in learning research: students do not construct meaning exclusively through verbal information. Diagrams, spatial arrangements, illustrations, photographs, maps, animations, symbols, gestures, and other representational resources can contribute to meaning-making. Multimodal learning therefore involves more than placing a picture beside a paragraph. It requires purposeful coordination between different modes so that each contributes appropriately to the learner's understanding.

This distinction is particularly important in the age of generative AI. A teacher can potentially create a visual representation of a difficult scientific process, a historical setting, a geographical phenomenon, a literary scene, or an abstract mathematical relationship without relying entirely on commercially available images. The same conceptual material can potentially be adapted for different levels of prior knowledge, languages, contexts, or learning activities. Recent scholarship describes this reduction in the production barrier as one of the major educational implications of generative AI.

However, increased production capacity does not automatically produce better learning. An attractive image may contain inaccurate information, irrelevant decoration, stereotypical representations, or excessive visual complexity. Generative systems can also produce apparently authoritative visualizations of concepts that require careful disciplinary verification. Thus, the central question is not whether AI can create educational images. It clearly can. The more important question is how such images should be designed, evaluated, integrated, and interpreted within pedagogically meaningful learning environments.

This article therefore investigates the following central question: **How can AI-generated visual content contribute to multimodal learning while preserving pedagogical quality, human judgment, accuracy, and learner agency?**

The article addresses four related objectives. First, it examines the relationship between AI-generated visuals and multimodal pedagogy. Second, it identifies potential educational applications of generated visual material. Third, it analyses limitations and ethical concerns surrounding AI-mediated visualization. Finally, it proposes a human-centred framework for incorporating AI-generated visuals into contemporary teaching and learning.

2. Literature Review

2.1 From Text-Centred Instruction to Multimodal Learning

Traditional educational communication has frequently privileged written and spoken language. Although images and diagrams have always existed in educational materials, they were often treated as supporting elements surrounding the principal textual explanation. Contemporary multimodal theory challenges this hierarchy by recognizing that meaning can be constructed through multiple communicative modes.

Multimodality considers language, image, sound, movement, spatial organization, and other semiotic resources as interconnected components of communication. In educational contexts, a lesson may therefore involve simultaneous interpretation of written instructions, diagrams, speech, gestures, animation, colour, and spatial relationships. The learner does not necessarily process these resources as isolated pieces; rather, understanding can emerge through their interaction.

This perspective provides an important foundation for understanding AI-generated visuals. If an AI-produced image is merely decorative, its educational contribution may be limited. If, however, it provides a spatial representation of a process that would otherwise be difficult to conceptualize verbally, it can become an integral component of meaning-making.

The shift toward multimodal education has also been accelerated by digital technologies. Interactive whiteboards, learning management systems, video platforms, visualization software, virtual reality, and mobile devices have made multimodal resources increasingly accessible. Generative AI extends this trajectory by making the production and modification of visual material considerably more accessible to ordinary users.

2.2 Multimedia Learning and Cognitive Processing

Research on multimedia learning provides another theoretical foundation for understanding AI-generated educational visuals. Mayer's cognitive theory of multimedia learning proposes that learners process information through interconnected verbal and visual channels and that meaningful learning requires active cognitive processing rather than passive exposure.

The implication for AI-generated images is significant. A visual should not be included merely because it looks impressive. It should help learners select relevant information, organize representations, and integrate those representations with prior knowledge.

For example, a generated diagram of the water cycle may support learning when it clearly represents evaporation, condensation, precipitation, and collection. A visually elaborate image containing numerous unrelated objects may instead increase cognitive demands without improving conceptual understanding.

Consequently, the pedagogical effectiveness of AI-generated visuals depends partly upon principles of coherence, signalling, spatial organization, and alignment between words and images. AI provides the means of producing visual content, but pedagogical theory determines whether that content is likely to be useful.

2.3 Generative AI and Educational Visualization

Recent research indicates growing scholarly attention toward the use of generative AI for visual learning. A 2025 systematic review examining AI-generated visual content in education identified

the visual dimension as an important but comparatively underexplored aspect of the broader generative-AI literature.

A 2026 survey of generative AI and three-dimensional visualization systems similarly found that AI content generation and immersive visualization are often developing along separate technological pathways, with relatively few systems successfully combining the two. The review also noted that teacher participation in reviewing or refining AI-generated educational content remains limited in many systems.

These observations suggest that the technological capacity to generate visual content has developed more quickly than the pedagogical frameworks necessary to govern its use. The challenge is therefore not simply technological innovation but educational integration.

2.4 Evidence from Language Learning

Recent empirical research provides an illustrative example. Ye and Yan's 2026 study compared text-only instruction with instruction incorporating AI-generated images in English vocabulary learning among 40 learners in China. The researchers reported stronger performance for the multimodal group on several measures of recall and semantic understanding. Importantly, however, the researchers also acknowledged that their design could not establish that AI-generated images themselves were responsible for the advantage because the study did not include a comparison condition using conventional images.

This limitation is theoretically important. It demonstrates why claims about AI-generated visuals must be separated from claims about multimodal instruction more generally. The evidence may show that combining words and relevant images supports learning, but this does not automatically mean that AI-generated images outperform teacher-created, textbook-based, photographic, or traditionally designed visuals.

2.5 Multimodal Composition and Learner Agency

AI-generated images can also alter the learner's role. Instead of merely consuming visual material created by teachers or publishers, students may use generative systems to translate ideas into visual representations. This creates possibilities for multimodal composition in which learners combine written explanations, images, diagrams, narration, and other modes.

Recent research on image-generative AI and digital multimodal composing has explored how learners can use generated images to express meanings that might be difficult to communicate through language alone. At the same time, researchers emphasize that the educational value of such practices must be considered alongside questions of learner agency, authorship, and critical engagement.

Thus, AI-generated visualization can be understood not only as a teacher resource but also as a potential medium for student expression.

3. Research Methodology

This study employs a **qualitative conceptual research design** based on critical analysis of contemporary scholarship concerning generative AI, multimodal learning, multimedia learning, and educational visualization.

The study uses secondary sources including peer-reviewed research articles, scholarly reviews, and international educational guidance. Particular attention is given to recent literature published during the development of contemporary generative-AI systems, while foundational concepts from multimedia and multimodal learning theory are used to establish the theoretical basis of the discussion.

The analysis follows four thematic categories:

1. **Pedagogical affordances** — the ways AI-generated visuals may support explanation, comprehension, engagement, personalization, and multimodal composition.
2. **Cognitive considerations** — the relationship between visual complexity, attention, information processing, and meaningful learning.
3. **Ethical and educational limitations** — including accuracy, bias, privacy, authorship, accessibility, and cultural representation.
4. **Teacher and learner agency** — the extent to which AI should function as an instructional assistant rather than an autonomous educational authority.

The study does not claim to provide experimental evidence of causal learning effects. Instead, it synthesizes available scholarship to develop a conceptual framework for responsible use of AI-generated visual content in multimodal education.

4. AI-Generated Visuals as an Extension of the Textbook

The phrase "beyond the textbook" should not be interpreted as a rejection of textbooks. Textbooks remain valuable because they provide structured, curated, and relatively stable bodies of knowledge. The emerging educational environment instead suggests that textbooks can become one component within a broader multimodal ecosystem.

AI-generated visuals can extend textbook content in several ways.

4.1 Visualizing Abstract Concepts

Many academic concepts are difficult to communicate exclusively through prose. Scientific processes, mathematical relationships, historical environments, geographical systems, architectural structures, and theoretical models can benefit from visual representation.

AI can potentially generate an initial visualization from a teacher's description. The teacher can then verify and modify the result before presenting it to students.

The process therefore changes from:

Textbook explanation → student interpretation

to a potentially richer sequence:

Textbook explanation → AI-assisted visualization → teacher verification → multimodal presentation → student interpretation

The AI-generated image does not replace disciplinary knowledge. It acts as a bridge between abstract information and perceptual representation.

4.2 Contextualizing Knowledge

Educational images are more meaningful when learners can connect them to familiar contexts. Generative systems can potentially create contextual variations of the same concept.

For instance, a lesson on architecture might use examples from different geographical regions. A language lesson might place vocabulary within culturally recognizable situations. A geography lesson might represent environmental processes within different landscapes.

Such flexibility can support contextualized instruction, provided that generated representations are culturally responsible and factually checked.

4.3 Supporting Differentiated Instruction

Learners vary in prior knowledge, language proficiency, and preferred ways of engaging with information. AI-assisted visual generation may allow educators to produce multiple versions of the same conceptual material.

A teacher might create:

- a simple visual for introductory learners;
- a labelled diagram for intermediate learners;
- a detailed technical representation for advanced learners;
- a bilingual visual for language support;
- an interactive or sequential visual for process-based learning.

This possibility aligns with broader research suggesting that multimodal AI systems can adapt educational content across modalities and potentially support diverse learners.

4.4 Supporting Multimodal Student Production

AI-generated visuals can also move learning from reception toward production. Students can formulate prompts, evaluate generated images, identify inaccuracies, revise visual representations, and combine them with their own written or spoken explanations.

Such activities can cultivate a form of visual literacy in which learners ask not merely "What does this picture show?" but also:

- Who produced it?
- What assumptions does it contain?
- Which information is emphasized?
- What information is missing?
- Is the representation accurate?
- How does the image influence interpretation?

This critical dimension is particularly important because generative AI can produce persuasive-looking representations that nevertheless require verification.

5. Educational Applications

5.1 Science Education

Science education provides a particularly strong context for visualization. Processes such as cellular division, planetary movement, molecular interaction, ecological systems, and geological change often involve spatial or temporal relationships that are difficult to represent through prose alone.

AI-generated diagrams or sequential illustrations can provide supplementary representations. However, scientific accuracy must remain the responsibility of the teacher and subject specialist.

5.2 Language and Literature Education

In language education, AI-generated images can support vocabulary acquisition, contextual learning, narrative interpretation, and multimodal composition.

Literature classes may use visual representations as prompts for discussing setting, symbolism, character, atmosphere, or historical context. Students could compare a textual description with an AI-generated visualization and critically examine what the image adds, omits, or changes.

Such an approach can turn visual generation into an interpretive activity rather than a purely technological exercise.

5.3 History and Social Studies

Historical education presents both opportunities and substantial risks. AI-generated reconstructions can help students visualize architecture, clothing, environments, or everyday settings associated with particular historical periods.

Nevertheless, reconstruction is not equivalent to historical evidence. A generated image should be explicitly identified as an interpretation or reconstruction rather than presented as a documentary record.

This distinction can itself become a valuable lesson in historical methodology.

5.4 Mathematics

Mathematical concepts can benefit from generated diagrams, coordinate representations, geometric models, and visual explanations.

The principal value lies not in aesthetic presentation but in making relationships visible. A generated visual can assist learners in connecting symbolic expressions with spatial or graphical representations.

5.5 Teacher Education

Teacher-training programs can use AI-generated visuals to explore lesson design. Trainee teachers may be asked to generate an image, evaluate it for pedagogical quality, identify inaccuracies, revise it, and justify their final version.

This makes AI literacy part of pedagogical literacy.

6. Challenges and Limitations

6.1 Accuracy and Hallucination

One of the most important problems is factual reliability. Generative AI systems may create visually convincing but incorrect objects, structures, labels, relationships, or historical representations.

In education, visual plausibility can be particularly dangerous because learners may assume that an apparently professional image is authoritative.

Every AI-generated educational visual should therefore undergo human verification before instructional use.

6.2 Cognitive Overload

More visual information does not necessarily mean more learning. Excessive decoration, unnecessary animation, complicated backgrounds, and dense visual compositions may compete for learners' attention.

The objective should be **instructional clarity rather than visual abundance**.
This principle suggests that educators should evaluate generated images according to what they help learners understand, not simply according to their aesthetic quality.

6.3 Bias and Representation

AI systems learn patterns from large datasets that contain cultural assumptions and representational imbalances. Generated educational visuals may therefore reproduce stereotypes or disproportionately represent particular cultures, regions, genders, occupations, or social groups.
UNESCO emphasizes the importance of equity, inclusion, cultural diversity, linguistic diversity, and human agency in educational uses of generative AI.
Consequently, teachers should critically inspect generated representations rather than assuming neutrality.

6.4 Authorship and Academic Integrity

The use of generated visuals raises questions about authorship. If a student submits an AI-generated diagram as part of an assignment, the educational significance depends partly on what the student actually contributed.
Educational institutions therefore need transparent expectations concerning disclosure, attribution, modification, and responsible use.

6.5 Privacy

AI-based educational tools may process prompts, uploaded images, student information, or classroom materials. UNESCO has specifically identified data privacy as an important concern in the educational use of generative AI.
Institutions should therefore establish clear rules regarding what information can be entered into AI systems.

6.6 Teacher Displacement Concerns

The ability to generate lesson visuals rapidly may create the misconception that instructional expertise can be automated. Yet producing an image and knowing why, when, and how to use it are different skills.
Recent scholarship argues that the increased availability of generative tools makes pedagogical expertise more important rather than less important.
The teacher remains responsible for interpreting curriculum objectives, assessing learner needs, checking accuracy, selecting appropriate resources, and facilitating discussion.

7. Discussion

The central finding of this conceptual analysis is that AI-generated visuals should be understood as **pedagogical instruments rather than pedagogical substitutes**.
The distinction matters because technology can automate production without automating educational judgment. An AI system may generate twenty possible diagrams in seconds, but it cannot independently determine which representation is most appropriate for a particular group of learners, curricular objective, cultural setting, or assessment task.
Recent research supports a cautious interpretation of the emerging evidence. The 2026 PLOS ONE study by Ye and Yan found positive learning outcomes when AI-generated images were

incorporated into vocabulary instruction, but the researchers explicitly noted that the study could not isolate the effect of AI-generated imagery from the broader effect of multimodal instruction. This distinction prevents technological enthusiasm from becoming an unsupported educational claim.

The future of multimodal learning may therefore be conceptualized as a **human-AI-learner relationship**:

Teacher expertise + AI generation/adaptation + learner interpretation = multimodal learning environment

Within this model, AI performs productive functions such as generating alternatives, adapting formats, visualizing concepts, and assisting with resource development. Teachers perform evaluative and pedagogical functions. Learners remain active interpreters and producers of knowledge.

7.1 A Proposed Human-Centred Visual Learning Framework

The article proposes a six-stage framework:

Stage 1: Learning Objective

The teacher identifies precisely what learners should understand or accomplish.

Stage 2: Visual Need

The teacher determines whether a visual representation genuinely contributes to the objective.

Stage 3: AI-Assisted Generation

AI is used to produce one or more possible visual representations.

Stage 4: Human Verification

The teacher checks factual accuracy, cultural representation, accessibility, cognitive complexity, and curricular relevance.

Stage 5: Multimodal Integration

The verified visual is combined with appropriate text, speech, interaction, discussion, or activity.

Stage 6: Critical Learner Engagement

Students interpret, question, compare, modify, or produce visual representations themselves.

This framework changes the role of AI from **content authority** to **content assistant**.

8. Implications for Educators

Teachers adopting AI-generated visuals should consider five practical principles.

First, every visual should have a clearly defined instructional purpose.

Second, generated material should be checked before classroom use.

Third, students should be taught to question AI-generated representations rather than accepting them automatically.

Fourth, visual materials should be designed with accessibility and cultural diversity in mind.
Fifth, AI use should be disclosed when it meaningfully contributes to educational products or assignments.
These principles align with the human-centred orientation recommended by UNESCO, which emphasizes that generative AI should be integrated in ways that protect human agency and promote safe, equitable, meaningful educational practice.

9. Conclusion

The movement beyond the textbook does not necessarily represent the disappearance of the textbook. Rather, it represents a transition toward educational environments in which written knowledge interacts with visual, auditory, spatial, and interactive forms of representation.
AI-generated visuals have considerable potential within this environment. They can help educators visualize difficult concepts, adapt materials, contextualize instruction, support multimodal composition, and create alternative representations of knowledge. Recent research provides encouraging evidence that multimodal instruction incorporating AI-generated images can support certain learning outcomes, although the existing evidence remains insufficient to conclude that AI-generated imagery is inherently superior to conventional visual resources.
The major challenge is therefore not whether educational institutions should use AI-generated visuals, but how they should use them responsibly.
A productive future for AI-enhanced visual learning requires a shift away from technological novelty toward pedagogical purpose. Images should be generated because they help learners understand, question, compare, create, or remember—not simply because artificial intelligence makes their production easy.
The most valuable educational model is consequently neither textbook-only instruction nor AI-dominated instruction. It is a multimodal learning environment in which established academic knowledge, teacher expertise, generative technology, visual literacy, and active learner interpretation operate together.

References

Istrate, O. (2026). Generative AI and multimodal pedagogy: Implications for teaching, learning, and assessment. *Frontiers in Education, 11*. https://doi.org/10.3389/feduc.2026.1828953

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Vilaça, P., & Faria, P. M. (2026). Capabilities of generative artificial intelligence and three-dimensional visualizations in education: A survey of recent systems. *Computer Applications in Engineering Education*. https://doi.org/10.1002/cae.70243

Ye, G., & Yan, S. (2026). Multimodal instruction with AI-generated images for noun retention: Exploring semantic scene and materiality effects. *PLOS ONE, 21*(4), e0334778. https://doi.org/10.1371/journal.pone.0334778

Wang, D., Wei, S., Chen, X., Chai, C.-S., Chen, G., & Hu, L. (2025). *AI-generated visual content in education: A systematic review of the past decade*. SSRN. https://doi.org/10.2139/ssrn.5935496