GAINS: A Generative AI and Augmented Reality Framework for Multimodal Storytelling in Education

Masood Azhar1, a), Fareha Nizam2, b) and Hew Soon Hin1, 3, c)

1Faculty of Creative Multimedia, Multimedia University, Cyberjaya, 63100, Selangor, Malaysia.

2Faculty of Artificial Intelligence and Engineering, Multimedia University, Cyberjaya, 63100, Selangor, Malaysia.

3Centre for Innovative and Immersive Technology, COE For Immersive Experience, Multimedia University, Cyberjaya, 63100, Selangor, Malaysia.

c) Corresponding author: shhew@mmu.edu.my

a) masoodazharpk@gmail.com

b) farehanizam2@gmail.com

**Abstract.** Generative artificial intelligence (generative AI) and augmented reality (AR) are transforming education by enabling multimodal, interactive storytelling experiences. While generative AI creates personalized, dynamic narratives, AR enhances engagement through real-time visualizations. However, integrating these technologies into structured educational models presents challenges in narrative coherence, multimodal synchronization, and pedagogical alignment. This paper introduces GAINS (Generative AI for immersive narrative storytelling), a conceptual framework that combines AI-generated storytelling, augmented interactivity, and adaptive learning strategies. We explore technical and design challenges in AI-AR storytelling, including context inconsistencies, personalization limitations, and ethical concerns in AI-generated narratives. Finally, a structured GAINS framework is proposed to align interactive narratives with augmented reality, offering a basis for the future development of advanced storytelling tools in education.

# Introduction

The practice of storytelling in human interaction and education has always served a basic, integral function, allowing for the transmission of knowledge, culture, wisdom, and life experiences over multiple generations [1]. In formal education, story-based teaching and learning practices improve engagement, aid in remembering important details, and help explain complex concepts in a more tangible way [2]. Innovation in education, particularly the shift to a more learner-centered model, has made it possible to combine storytelling with modern technologies, creating new avenues for tailored and more engaging styles of learning [3].

Recent advances in generative artificial intelligence, specifically large language models (LLMs), facilitate the automated synthesis of structured texts, stories, and other forms of writing. These systems are capable of performing tasks ranging from conversation emulation to story progression and even narrative modulation to encompass emotional or didactic elements [4]. Such capabilities are changing the traditional model of the educational system, allowing the learners to take on the role of active and dynamic users in the evolving educational story, not passive viewers of unchanging materials [5].

The technology of augmented reality (AR) has concurrently emerged as a robust platform for multimodal educational visualization [6]. Consider molecular models, historical dates, and even advanced mathematical functions; with AR, users can interact with three-dimensional and spatial representations interactively, allowing learners to grasp abstract concepts [7, 8]. Coupled with adaptive storytelling, AR provides a synergistic ecosystem for acquisition by allowing learners to traverse through rich, immersive narratives while manipulating contextually embedded visuals within their environment [9].

Nonetheless, the application of such a technological integration poses considerable difficulties. Constructing real-time narratives that are relevant and pedagogically appropriate requires the integration of narrative design, learning objectives, and the learner’s profile. Augmented reality content, if provided without appropriate context or alignment with the overarching narrative, risks disengagement and cognitive disinterest. Also, the fluidity of AI-AR interactions may be hindered by network, hardware, and device compatibility gaps. Gaps of enabling differently abled users involve design changes, such as adaptive interaction methods, tailored narrative intricacy, and adjustable engagement levels. In addition, employing such technologies raises pertinent concerns related to the algorithms used in content generation, including biases within the algorithms, privacy of user data, cultural sensitivity, and the overall trustworthiness of the content generated. In the absence of sound design principles, educational applications of such technologies are likely to result in disjointed and ineffective, or even harmful, learning experiences.

These intricacies emphasize the necessity of having a consistent model that Augmented Reality (AR) and generative AI integration into pedagogically sound educational storytelling systems. Many educational technologies still approach these elements as separate problems, algorithmically processing and generating the content, and visualizing affordances pedagogically, and seldom treating both as equally important. Moreover, although there are numerous applications of AR in STEM education, and some have attempted integration of narrative learning, there are limited solutions that address the fully interactive, adaptive storytelling model in AR-based educational systems.

In this paper, we explore the GAINS (Generative AI for Immersive Narrative Storytelling) framework, which weaves together elements of generative AI, adaptive learning, augmented reality (AR), and pedagogical theory into a cohesive, multi-layered, and instructional design-oriented ecosystem. Within the GAINS framework, generative context-aware educational technologies grounded in ethical principles and learner-centered responsiveness are sought. We also inquire into the aims of widening the scope of primary multi-disciplinary adopted GAINS for STEM, humanities, and vocational training. With the intent of encouraging GAINS to be usable for many educational goals, it also aims to foster marked modularity to enable the framework to respond to various curriculum needs like collaborative, inquiry-driven, and simulation-based learning. The framework describes feedback and AI-driven narrative and immersive AR learner-content interactions, and addresses concerns like narrative coherence, multimodal interplay, and personalization. Moreover, the framework elaborates on the decisive elements of technical infrastructure, instructional design frameworks, and ethical boundaries that are necessary for GAINS to be deployed actively and meaningfully on a large scale.

This study introduces the GAINS framework as a structured approach to guide the development of next-generation educational technologies. It provides practical insights for system developers, curriculum designers, and researchers aiming to integrate AR and generative systems in ways that are both pedagogically effective and ethically informed. This approach addresses a critical gap in the literature by presenting a unified, theory-informed model that prioritizes learner engagement, cognitive development, and inclusivity.

# RELATED WORK

## Multimodal Learning and Storytelling

Multimodal learning draws on multiple sensory modalities, including visual, auditory, textual, and kinesthetic inputs, to enhance instructional effectiveness. Mayer’s Cognitive Theory of Multimedia Learning [10],[11] highlights that learners comprehend and remember information better when it is offered through more than one channel. The theory applies to the development of educational software that includes graphics, animation, and narration, as well as interactive elements that require learners to process information cognitively.

Storytelling as a technique strengthens multimodal learning by situating information within emotional and coherent narratives. Digital storytelling fosters critical thinking, motivation, and knowledge construction by integrating narrative with multimedia components [3]. In this case, storytelling goes beyond a delivery technique; it serves as a cognitive scaffold that bolsters understanding and engagement that is sustained over time. Narrative-based multimodal approaches have the potential to help learners understand the gap between abstract theoretical frameworks and concrete applications, especially when learners are prompted to actively and interactively manipulate digital objects within a story.

## Generative AI in Education

The advent of large language models, such as GPT-4, has changed the landscape of education personalization. With minimal prompts, these models are capable of producing contextually relevant and high-quality content, such as quizzes, feedback, and even adaptive narratives [4,5]. The application of generative AI has made it possible to support adaptive tutoring, simulate historical dialogues, and even assist in creative writing.

Khanmigo [12] and similar educational platforms are now incorporating LLMs for life-long learning and dynamic, personalized answers to user queries. Despite this, posture and voice interactivity through immersive or generative technologies is mostly absent in educational chatbots, limiting learner engagement through hands-on and exploratory activities. Integrating LLMs with AR systems has the potential to move beyond static or purely conversational modes, enabling context-aware content that dynamically responds to both the learner’s input and their physical surroundings. This approach can also foster collaborative problem-solving where multiple learners engage with a shared AI-AR narrative space.

## Augmented Reality for Learning

AR superimposes digital content onto the physical environment, creating hybrid learning spaces that support spatial interaction and visualization. AR has proven effective in domains requiring spatial reasoning and embodied cognition, such as anatomy, engineering, and language learning [7].

AR excels at contextualizing abstract concepts. For instance, AR anatomy tools allow users to explore body systems in 3D, while AR language apps situate vocabulary in real-world contexts [8]. Meta-analyses indicate that AR enhances student engagement, motivation, and conceptual understanding when aligned with instructional objectives. However, many AR applications rely on static content and lack the adaptive, personalized capabilities offered by AI [9]. Emerging research on AR learning environments suggests that adaptability, narrative integration, and learner agency are critical for sustained engagement, and these are qualities that generative AI can directly enhance.

## Integrated AI-AR Systems

Despite the promise of generative AI and AR as individual technologies, their integration into unified educational systems remains nascent [13]. A recent experimental prototype called SPARC demonstrated the combination of LLM-generated dialogue with AR visualizations, embedding real-time AI-curated content into immersive environments [14]. While the results highlight the feasibility of such integration, the study also identified important next steps, including improving NLP recognition for children’s speech patterns and nonstandard question structures, expanding multilingual support, and conducting large-scale usability testing to evaluate long-term learning outcomes.

Moreover, many integrated systems lack robust pedagogical frameworks. Without instructional coherence, they risk delivering visually rich but educationally shallow experiences. Current literature does not yet offer a comprehensive model that bridges generative AI’s generative capabilities with AR’s embodied affordances while maintaining narrative continuity, learner agency, and ethical safeguards.

A significant gap in current integrated prototypes is the lack of robust mechanisms for real-time, bidirectional influence between the generative narrative and the AR visualization. Often, the AR elements are pre-scripted or only loosely triggered by the narrative, rather than being dynamically generated or modified by the AI in direct response to both learner actions within the AR space and the evolving story. Achieving this level of tight, context-sensitive multimodal coupling, where changes in the narrative immediately and meaningfully alter the AR environment and vice versa, remains a complex technical and design challenge. The GAINS framework is positioned to address this gap by proposing a structured, theory-informed integration model, offering a foundation for future empirical testing and large-scale deployment.

## Motivation for the Proposed Framework

Despite significant progress in generative AI and augmented reality for education, current implementations often remain fragmented. AI-powered systems excel at producing adaptive, context-sensitive narratives but typically lack spatial immersion, while AR applications provide rich visualizations yet rely on static, pre-scripted content that fails to adapt to learner input. Such a split results in a disjointed experience in which stories, contexts, and assessments fail to align. Without clear learning goals or guiding principles, many blended designs run the potential risk of looking visually eye-catching, yet pedagogically empty, offering participation with little learning value. This is even more important in the STEM domains, which require learners to be meaningfully interested, to have a contextual understanding of important ideas, and to work together to develop understanding.

Generative AI, augmented reality, and Bandura’s social learning theory convergence is systematically elaborated within the GAINS framework, which, in its model, combines adaptive, real-time responsive narratives with immersively visualized narratives. GAINS combines the principles of observational learning and collaborative participation, transforming AR from static visualization tools to dynamic interactive spaces which learners engage with actively, thereby enhancing the development of their agency and identity. This model also actively and purposefully incorporates technology, ensuring that educators do not teach in isolation, thereby transforming instruction and learning in all subjects while taking responsibility for the impact on the learners’ achievement outcomes.

# PROPOSED FRAMEWORK

By combining augmented reality, generative AI, and Bandura’s social learning theory, the GAINS framework facilitates the adaptive and immersive learning experience in STEM education. The framework functions on three layers: Contextual Fusion, Interpretive Engine, and Transformational Impact, as seen in Figure 1 and described in Figure 2. It merges teaching methods with technology alongside deep learning and active engagement. Primarily centered on STEM education, the design principles, however, transcend temporal and spatial limits and can thus be applied to the humanities, vocational education, and interdisciplinary curricula.

figure 1.drawio

**FIGURE 1.** Layered design of AR-GAI enhanced learning

## Contextual Fusion Layer (Input Layer)

This layer combines key technologies and learning principles to personalize the learning context.

1. AR Integration: AR overlays interactive 3D visualizations, such as graphs and simulations, onto the physical environment, enabling learners to engage with abstract concepts through embodied interaction and real-time feedback.
2. Generative AI Personalization: Generative AI tailors content based on user performance, adjusting difficulty levels and generating contextual explanations, questions, or stories that align with each learner’s cognitive abilities and progress.
3. Social Learning Theory: Based on Bandura’s principles of observational learning and modeling, this layer supports active learning and critical thinking by allowing students to observe, imitate, and interact with intelligent agents or peers in AR environments.
4. Learner Profiling: The system uses individual learner data, including prior knowledge, behavior, and preferences, to inform its personalization engine, ensuring content delivery is both relevant and appropriately challenging.

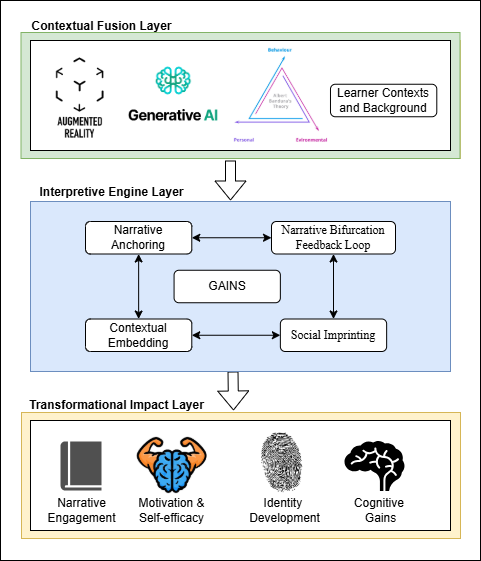
These parts integrate to facilitate responsive, learner-focused frameworks that enable narrative-driven immersion. Crucially, the contextual fusion layer contributes to ensuring that personalization goes beyond the scope of the content’s difficulty, incorporating the narrative’s style, pacing, and visuals. This allows differentiated instruction to be implemented at scale.

## Interpretive Engine Layer (Processing)

This layer synthesizes narrative, feedback, and spatial interaction to form a unified learning experience.

1. Narrative Anchoring: Learners interact with personalized narratives tailored to their cultural, emotional, and academic contexts, fostering engagement based on identity and emotional connection.
2. Contextual Embedding: Augmented reality situates learning within both spatial and temporal contexts, linking abstract knowledge to concrete experiences. Learners engage physically with content, supporting embodied cognition.
3. Narrative Bifurcation & Feedback Loop: Learners shape the storyline through decisions and dialogue, transitioning from passive recipients to active participants. Tailored feedback is provided, encouraging metacognitive reflection and deeper learning.
4. Social Imprinting: Collaborative tasks, such as co-authoring stories or participating in shared augmented reality experiences, promote social learning and meaning-making, extending the learning process beyond individual cognition.

Meeting the requirements of both personalized instruction and wide-scale implementation, the interpretive engine upholds narrative coherence by managing user interactions in real time, which enables adaptive branching pathways.



**FIGURE 2**. Proposed GAINS framework for immersive and personalized learning

## Transformational Impact Layer (Output)

The final layer assesses and enhances learning outcomes through four key impact areas:

1. Narrative Engagement: Learners stay actively engaged through immersive storytelling, facilitating deep interaction with the content.
2. Motivation & Self-Efficacy: The personalized and adaptive nature of narratives, enhanced by generative AI and AR, boosts learner confidence and intrinsic motivation.
3. Identity Development: Through engagement with culturally and socially relevant narratives, learners explore their identities, reinforcing both personal and academic self-concept.
4. Cognitive Gains: The integration of multimodal learning, problem-solving, and reflective feedback promotes deeper understanding, critical thinking, and long-term knowledge retention.

The transformational impact layer additionally functions as a feedback system for trainers by giving insightful analytics on engagement patterns, understanding metrics, and collaboration competencies of the learners. This data can be leveraged to refine instructional design and improve AI-AR narrative models over time.

# CONCLUSION

This paper presented GAINS, a multilayered framework incorporating generative AI with augmented reality for storytelling in education that is immersive, adaptive, and pedagogically grounded. GAINS resolves core issues of narrative coherence, multimodal integration, learner tailoring, and personalization through contextual fusion and transformational impact interpretive layers. It integrates technological proficiency with pedagogical principles, especially Bandura’s social learning theory, to facilitate agile and adaptive learning.

The framework prioritizes not only technical integration but also ethics, culture, and inclusivity. These considerations are essential in dealing with inequity and disparity for the economically and socio-culturally diverse education systems from better equipped city schools to the neglected urban and rural educational centers. Also, the wide range of subjects from STEM to the humanities and interdisciplinary courses allows flexible blending of the framework. In addition, the framework is applicable to formal, informal, and hybrid learning environments. Finally, the framework’s modularity allows for stepwise adoption, enabling institutions to incorporate AI and AR technologies in alignment with their educational and technological readiness.

Based on this framework, GAINS offers a development model for next generation technologies, especially those that combine narrative driven engagement with embodied interactivity. This opens opportunities for deeper study and development in the realm of AI integrated into intelligent learning systems. To this end, GAINS should be refined starting from conceptual validation through pilot studies, core curriculum integration across various subjects, and the inclusion of new modalities like haptics and mixed reality which would enhance the framework’s scope.

ACknowledgments

This research has been funded by the Ministry of Higher Education Fundamental Research Grant Scheme (FRGS) with Grant No. FRGS/1/2022/SSI0/MMU/01/1The authors wish to acknowledge the students and faculty members of the Faculty of Creative Multimedia, Multimedia University, for their assistance in this research study.

# References

1. E. E. Peterson and K. M. Langellier, “Communication as storytelling,” in *Communication As…: Perspectives on Theory* (SAGE Publications, Inc., Thousand Oaks, CA, 2006), pp. 123–131.
2. R. L. Reyes and J. A. Villanueva, “Narrative-based concept representations: Fostering visual cognition in the introductory chemistry classroom,” J. Chem. Educ. **101**(3), 1106–1119 (2024).
3. F. Javed, ed., *Innovative Pedagogies for a Digital World,* in *Innovation Management for a Resilient Digital Economy* (IGI Global, 2025), pp. 231–260.
4. G. Ocak and B. Olur, “Creative writing and storytelling: Investigating how generative artificial intelligence technologies can aid in the creative writing process, encouraging students to develop their storytelling abilities,” in *Generative Artificial Intelligence in Higher Education: A Handbook for Educational Leaders,* edited by K. Pushpanadham, N. Nordin, and H. Norman (Asia-Europe Meeting Research Network, 2024), p. 95.
5. O. R. Román Jiménez, V. Saavedra Gastélum, C. Zubieta Ramírez, and C.A. González Almaguer, “Application of active learning techniques supported by generative AI in education,” in *Proceedings of the International Conference* (Seville, Spain, 2024), pp. 10408–10417.
6. J. Challenor and M. Ma, “A review of augmented reality applications for history education and heritage visualisation,” MTI **3**(2), 39 (2019).
7. S. Câmara Olim, V. Nisi, and T. Romão, “Augmented reality interactive experiences for multi-level chemistry understanding,” International Journal of Child-Computer Interaction **42**, 100681 (2024).
8. G. Singh, G. Singh, N. Tuli, and A. Mantri, “Hyperspace AR: An augmented reality application to enhance spatial skills and conceptual knowledge of students in trigonometry,” Multimed. Tools Appl. **83**(21), 60881–60902 (2023).
9. X. Wei, L. Wang, L. -K. Lee, and R. Liu, “Multiple generative AI pedagogical agents in augmented reality environments: A study on implementing the 5E model in science education,” Journal of Educational Computing Research **63**(2), 336–371 (2025).
10. M. Azhar, H.S. Hin, and N.T. Kian, “SPARC: A framework for enhancing early childhood reading through augmented reality, conversational AI, and IDC theory,” in *Proceedings of the 2024 International Conference on TVET Excellence & Development (ICTeD)* (IEEE, Melaka, Malaysia, 2024), pp. 103–106.
11. R.E. Mayer, *The Cambridge Handbook of Multimedia Learning,* 1st ed. (Cambridge Univ. Press, Cambridge, 2012).
12. S. Shetye, “An evaluation of Khanmigo, a generative AI tool, as a computer-assisted language learning app,” SALT **24**(1), (2024).
13. Y. Kuerban, S.S. Oyelere, and I.T. Sanusi, “ReadSmart: Generative AI and augmented reality solution for supporting students with dyslexia learning disabilities,” IJTES **9**(1), 159–176 (2025).
14. M. Azhar, H. S. Hin, and N. T. Kian, “Enhancing early childhood reading skills through immersive AR and conversational AI,” in *Proceedings of the 2025 13th International Conference on Information and Education Technology (ICIET)* (IEEE, Fukuyama, Japan, 2025), pp. 71–75