
International Journal of Creative Multimedia

Narrative Co-Construction with Conversational AI in Augmented Reality: A Conceptual Framework for Collaborative Story-Making

Masood Azhar
masood.azhar@student.mmu.edu.my
Multimedia University, Malaysia

Soon Hin Hew
shhew@mmu.edu.my
Multimedia University, Malaysia
ORCID iD: 0000-0001-8355-3329
(Corresponding Author)

Tse-Kian Neo
tkneo@mmu.edu.my
Multimedia University, Malaysia
ORCIDiD:0000-0002-5991-7409

Abstract

Augmented reality (AR) and conversational artificial intelligence (ConvAI) have been explored as design components in interactive learning environments for children, particularly in literacy and storytelling contexts. However, most existing systems position artificial intelligence (AI) as a narrator, with learners primarily acting as recipients of pre-authored content. This interaction model limits learner agency and constrains opportunities for shared meaning-making. To address this gap, this paper introduces narrative co-construction, in which children and AI collaboratively create and shape stories through ongoing interaction. Narrative co-construction is conceptualised as a shift from storytelling to story-making, emphasising shared control, turn-taking, and dynamic narrative development within AR learning environments. Building on interaction design and child-centred learning principles, we propose a conceptual framework, the Collaborative Narrative Interaction Framework (CNIF), which outlines key components of co-constructive narrative interaction, including narrative initiative, conversational responsiveness, and adaptive story branching. The framework is illustrated through AR-based learning scenarios that demonstrate how ConvAI can function as a collaborative partner rather than a narrative authority. By reframing the role of AI from content provider to co-author, this work contributes design-oriented insights for developing AR learning environments that support creativity, engagement, and active meaning-making in children's learning experiences.

Keywords Augmented reality; Conversational artificial intelligence; Narrative co-construction; Collaborative narrative interaction; Learning design

Received: 22 June 2026, **Accepted:** 24 August 2026, **Published:** 30 September 2026.

Introduction

The development of early literacy skills constitutes the foundational bedrock of lifelong learning, critical thinking, and cognitive growth (Mol & Bus, 2011; Strickland & Riley-Ayers, 2006). Historically, early childhood education has relied upon structured, unidirectional methodologies, such as printed storybooks, rote phonics drills, and standardised curricula (Aligned, n.d.; Copple & Bredekamp, 2009). While these methods possess enduring value, they frequently fail to cater to the diverse developmental, emotional, and cognitive needs of individual learners, leading to inconsistent outcomes in reading comprehension and vocabulary development (Lerkkanen et al., 2016; Marcon, 2002; Stipek, 2006). The contemporary transition towards learner-centred education, augmented by advanced digital technologies, has introduced transformative opportunities to reimagine personalised and immersive learning environments. Among these technological innovations, augmented reality (AR) and conversational artificial intelligence (ConvAI) hold immense potential to radically reshape interactive educational experiences (Monjoree et al., 2024; Wu et al., 2026). AR superimposes interactive digital elements onto the physical environment, enabling multi-sensory, spatial exploration, while ConvAI simulates sophisticated human dialogue, dynamically adapting responses based on natural language processing (NLP) (Azuma, 1997; McTear, 2022).

Aligned with the ethos of reimagining creativity through artificial intelligence (AI), it is imperative to critically examine how AI can transcend its traditional utility as a mere instructional tool to become a collaborative creative agent. Despite their profound individual strengths, the integration of AR and ConvAI in early literacy applications has largely adhered to a conservative, traditional paradigm. Most existing systems position the AI as a unidirectional narrator, an authoritative tutor, or a static content provider, while young learners are relegated to the role of passive recipients consuming pre-authored material (Darvishi et al., 2024; Mai, 2025). Even in interactive digital storytelling environments, children's agency is frequently limited to selecting from pre-defined, algorithmic decision trees, effectively operating as sophisticated "pick-a-path" experiences rather than authentic creative endeavors (Robertson & Good, 2005; Zhang et al., 2022; Liu et al., 2025). This unidirectional interaction model fundamentally limits learner agency, constrains opportunities for shared meaning-making, and severely curtails the cognitive scaffolding necessary for deeper learning.

To address this critical pedagogical and technological gap, this paper conceptualises a paradigm shift from traditional "storytelling" to interactive "story-making". By introducing the concept of narrative co-construction, this research posits that children and AI must collaboratively create and shape stories through continuous, bidirectional, and symmetrical interaction. Reframing the role of AI from a narrative authority to an active co-author requires a foundational redesign of both technological architectures and instructional strategies. Building upon interaction design, advanced computer-

supported collaborative learning (CSCL), and child-centred pedagogical principles, this paper proposes the collaborative narrative interaction framework (CNIF). This conceptual framework operationalises shared control, conversational responsiveness, and adaptive story branching within immersive AR environments.

By examining the intersection of emergent agentic AI, multi-modal spatial computing, and pedagogical theories, such as the interest-driven creator (IDC) theory and the cognitive theory of multimedia learning (CTML), this paper explores the theoretical, architectural, and ethical dimensions of deploying co-constructive AI systems. A set of design principles for AR-based learning environments is proposed to foster creativity, deeper cognitive engagement, and dynamic meaning-making in early childhood education, positioning emerging technologies as catalysts for authentic human-machine collaborative storytelling.

Research Design and Approach

To establish a rigorous and logically sound foundation for the proposed CNIF, this paper adopts a theory-synthesis and model-building conceptual research design. Rather than conducting an empirical intervention, the paper integrates complementary theoretical and research perspectives to develop an explanatory framework for narrative co-construction in AR-ConvAI learning environments (Jaakkola, 2020). The synthesis draws on literature that directly informs the focal phenomenon from four complementary perspectives: CSCL for technologically mediated co-construction and shared meaning-making; the IDC theory for learner interest, creation, and sustained engagement; CTML for the cognitive processing of coordinated visual and verbal information; and the interactive, constructive, active, and passive (ICAP) framework, complemented by Vygotskian scaffolding, for explaining levels of learner engagement and adaptive support. These theoretical perspectives are examined for their complementary explanatory roles and translated into recurring design requirements for child-AI narrative collaboration. The synthesis culminates in three interconnected CNIF components, narrative initiative, conversational responsiveness, and adaptive story branching, which are subsequently illustrated through theoretical AR-based learning scenarios. Accordingly, the scenarios are intended as analytic illustrations of the framework's proposed operation rather than as empirical evidence of its effectiveness.

Theoretical Foundations of Narrative Co-Construction

The transition towards story-making in AR environments necessitates a robust theoretical underpinning that unifies developmental psychology, human-computer interaction, and instructional design. Educational technology deployed without a theoretical anchor risks becoming visually impressive but instructionally shallow. The CNIF is therefore built upon a synthesis of CSCL, the IDC theory, CTML,

and the interactive, constructive, active, and passive (ICAP) framework, all contextualised within the broader paradigms of social and observational learning.

Computer-Supported Collaborative Learning (CSCL)

CSCL provides the social-interactional foundation for the CNIF by framing learning as knowledge construction through technologically mediated interaction rather than as one-way information transmission. CSCL scholarship emphasises collaborative knowledge building, dialogic interaction, intersubjective meaning-making, and the emergence of understanding through interaction among participants (Stahl et al., 2006). More recent scholarship has begun to consider how selected principles of collaborative knowledge construction can inform human–AI interaction, while recognising that generative AI should not simply be equated with a human collaborator (Cress & Kimmerle, 2023). Within the CNIF, CSCL therefore functions as a design lens for reciprocity and uptake: narrative initiative enables the child to introduce, negotiate, and redirect ideas; conversational responsiveness requires the AI to acknowledge and meaningfully build upon those contributions; and adaptive story branching preserves the consequences of jointly developed decisions within the evolving narrative and AR environment. In this sense, collaboration within CNIF refers to an interaction design principle of reciprocal contribution and shared meaning-making rather than an assertion that the AI possesses human-like social understanding or agency.

Interest-Driven Creator (IDC) Theory

The IDC theory serves as a central pedagogical anchor for designing educational technologies that foster intrinsic motivation and sustained engagement, particularly in the Asian educational context and beyond (Chan et al., 2018). IDC theory operates on the premise that when learning experiences are propelled by personal curiosity, learners exhibit higher levels of creativity and persistence (Wong et al., 2020). The theory is structured around three iterative, interconnected loops: interest, creation, and habit (Chan et al., 2019).

The interest loop is designed to capture and maintain the learner's attention by aligning the educational stimulus with their intrinsic curiosities. In an AR/ConvAI context, this is achieved through visually stimulating spatial overlays and personalised narrative hooks that resonate with the child's existing knowledge or aesthetic preferences. Once interest is established, the learner transitions into the creation loop. Moving beyond passive observation, this loop demands active participation and knowledge generation. Narrative co-construction relies heavily on this phase; as the child engages in dialogue with the ConvAI, answers open-ended prompts, and dictates the trajectory of the storyline, they transition from a consumer of content to a creator of knowledge. Finally, to ensure long-term educational impact, the system must reinforce repeated engagement through the habit loop (Chen et al.,

2020). This loop consists of a cuing environment, a routine, and an affective outcome of harmony. By dynamically adapting to the child's evolving literacy level and providing an emotionally safe space for conversational experimentation, the system cultivates a sustained, habitual desire for reading and storytelling, ultimately fulfilling the learner's psychological needs and fostering an enduring identity as a creator.

Cognitive Theory of Multimedia Learning (CTML)

While the IDC theory drives motivation, the cognitive mechanics of processing multi-modal information within an immersive AR environment must be carefully managed. Mayer's (2021) CTML posits three core assumptions: humans possess dual channels for processing visual/pictorial and auditory/verbal information; these channels possess a strictly limited processing capacity; and meaningful learning only occurs through active cognitive processing, wherein learners select relevant information, organise it into coherent mental structures, and integrate it with prior knowledge.

Integrating AR and ConvAI presents a substantial risk of cognitive overload, particularly for young children whose executive functions (EFs) and working memory capacities are still maturing (Mayer, 2021; Diamond, 2013). CTML holds that instructional design should minimise extraneous processing, manage essential processing, and foster generative processing (Mayer, 2021). Therefore, within the proposed framework, AR visualisations must be designed with strict visual minimalism, utilising sequential reveal strategies to introduce story elements gradually. The ConvAI acts as the primary auditory/verbal channel, narrating or responding synchronously with the visual/pictorial channel of the AR interface. When these modalities are effectively synchronised without overwhelming the learner's sensory input, they reflect Mayer's (2021) multimedia principle that learners understand more deeply from coordinated words and pictures, such as narration paired with relevant visuals, than from words alone.

Interactive, Constructive, Active, and Passive (ICAP) Framework and Vygotskian Scaffolding

To understand the qualitative shift in learning outcomes produced by narrative co-construction, it is useful to examine the ICAP framework. Traditional AI storytelling applications relegate the user to a passive (receiving content) or active (tapping a screen to turn a page) role, which primarily supports "knowledge telling" (Chi & Wylie, 2014). Conversely, generative AI tools designed as collaborative writing partners facilitate constructive and interactive modes of learning. When children critically engage with AI output, revise it, and integrate it into their own imaginative ideas, the process shifts towards "knowledge transforming," leading to deeper cognitive embeddedness and more meaningful learning outcomes (Ojeda-Ramirez et al., 2025).

Furthermore, Bandura (1977) emphasises in his social learning theory that learning is a deeply social process occurring through observation, modelling, and collaborative interaction. When ConvAI is designed as an agentic peer rather than a didactic tutor, it models complex narrative structures, advanced vocabulary, and appropriate emotional expression. Drawing upon Vygotsky's (1978) concept of the zone of proximal development, the AI acts as a "more knowledgeable other," providing real-time, adaptive scaffolding. By utilising specific conversational strategies such as elaborating on the child's ideas, eliciting further details, and echoing correct phrasing during narrative co-construction, the AI companion challenges the child just enough to advance their linguistic and cognitive abilities without inducing the frustration that leads to disengagement.

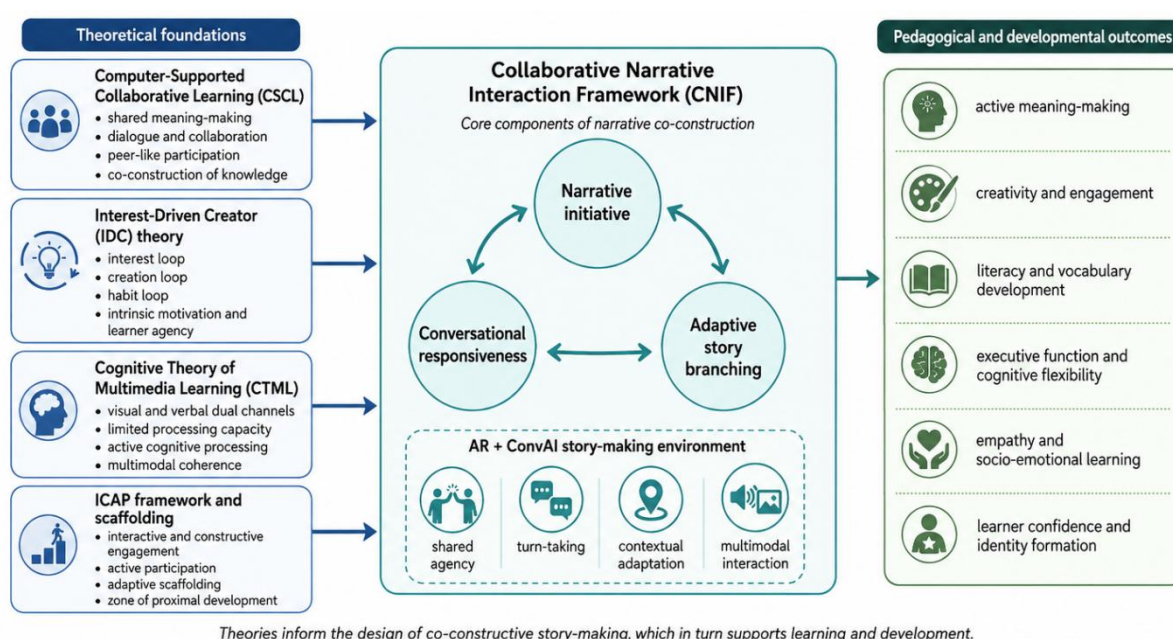


Figure 1. Theoretical Integration Underlying the Collaborative Narrative Interaction Framework (CNIF)

Figure 1 presents the four theoretical lenses as complementary foundations of the CNIF. CSCL informs reciprocal co-construction and shared meaning-making; IDC theory supports interest, creation, and learner agency; CTML guides the coordination of visual and verbal information; and the ICAP framework with scaffolding supports constructive, interactive, and adaptively scaffolded engagement. Together, these perspectives inform the three CNIF mechanisms: narrative initiative, conversational responsiveness, and adaptive story branching. The relationships shown are conceptual and require empirical validation.

Paradigm Shift: From Storytelling to Story-Making

The traditional paradigm of digital storytelling in educational technology predominantly features interactive multimedia where the narrative arc is predetermined. While AR applications have

successfully visualised abstract concepts, such as rendering a three-dimensional historical artefact or a fictional character in a child's physical environment, the underlying interaction model remains fundamentally one of consumption. In such environments, the child's agency is illusory; their interactions merely trigger pre-scripted events within a closed narrative loop.

Rise of Agentic AI

The reconceptualisation of this dynamic into "story-making" is made possible by the rapid evolution of agentic AI. Widely regarded as a major technological development of 2025 and 2026, agentic AI refers to systems capable of independent planning, reasoning, utilising tools, adapting to fluid conditions, and executing multi-step workflows with minimal human supervision (Mukherjee & Chang, 2025). Unlike purely reactive chatbots that wait for turn-by-turn instructions, agentic AI systems are oriented around ongoing task execution and contextual awareness.

Agentic AI transcends the identity of a mere cognitive tool to become a co-creator. Such an AI can articulate uncertainty, ask clarifying questions, learn from the child's specific imaginative tendencies, and dynamically integrate the child's contributions into a structurally sound, long-form narrative. This elevates the interaction from a transactional exchange to a genuine creative partnership, where the AI's proactive engagement is bounded, context-aware, and pedagogically purposeful.

Shared Agency and Turn-Taking Mechanics

A critical element of the story-making paradigm is the equitable distribution of control. In passive systems, algorithmic pacing dominates the user. Co-construction, however, requires "shared agency," an implicit, fluid arrangement where both human and machine readily adopt and contribute to one another's assertions, strategies, and narrative goals. This view of shared agency draws on distributed cognition, which treats a human–technology system as an integrated cognitive unit (Hutchins, 1995).

For early childhood learners, shared agency must be carefully mediated through structured conversational turn-taking. Research in child–computer interaction suggests that timely system responses and clear turn-taking cues are important for supporting children's successful interactions with conversational technologies (Beneteau et al., 2019; Lovato & Piper, 2015). If a conversational agent speaks continuously without allowing interruption, or fails to signal when it is listening, the collaborative illusion shatters. Shared control empowers the child, fostering a sense of psychological safety that is particularly crucial for marginalised or neurodivergent learners, allowing them to experiment with language and narrative without the fear of failure typically associated with hierarchical educational structures. Table 1 summarises this fundamental paradigm shift, contrasting the traditional dimensions of AI as a narrator against the proposed model of AI as an interactive co-author.

Table 1. The paradigm shift from ai storytelling to AI story-making

Paradigm Feature	Traditional AI Storytelling (AI as Narrator)	Interactive AI Story-Making (AI as Co-Author)
Primary Interaction Mode	Passive consumption / Reactive branching selection	Active co-creation / Proactive, generative contribution
Narrative Control	Centralised within a pre-authored, static algorithm	Distributed through shared agency, turn-taking, and mutual negotiation
Role of the AI	Didactic authority, content provider, and evaluator	Collaborative peer, creative partner, and adaptive cognitive scaffolder
Error Handling	Strictly corrects "wrong" answers to return to the script	Seamlessly incorporates unexpected or "illogical" inputs creatively into the plot
Underlying Technology	Static decision trees, basic NLP, predefined 3D assets	Agentic large language models (LLMs), real-time spatial computing, procedural asset rendering

The Collaborative Narrative Interaction Framework (CNIF)

To systematically implement the pedagogical and technological shift from storytelling to story-making, this paper introduces CNIF. This conceptual blueprint, expanding upon the foundational architectures of the system for personalized AR and reading with ConvAI (SPARC), outlines three interconnected components essential for achieving true co-constructive interaction within AR learning environments: narrative initiative, conversational responsiveness, and adaptive story branching (Azhar et al., 2024). As illustrated in Figure 2, this framework visually translates the shift from traditional storytelling to dynamic story-making by detailing the iterative loop required for shared meaning-making.

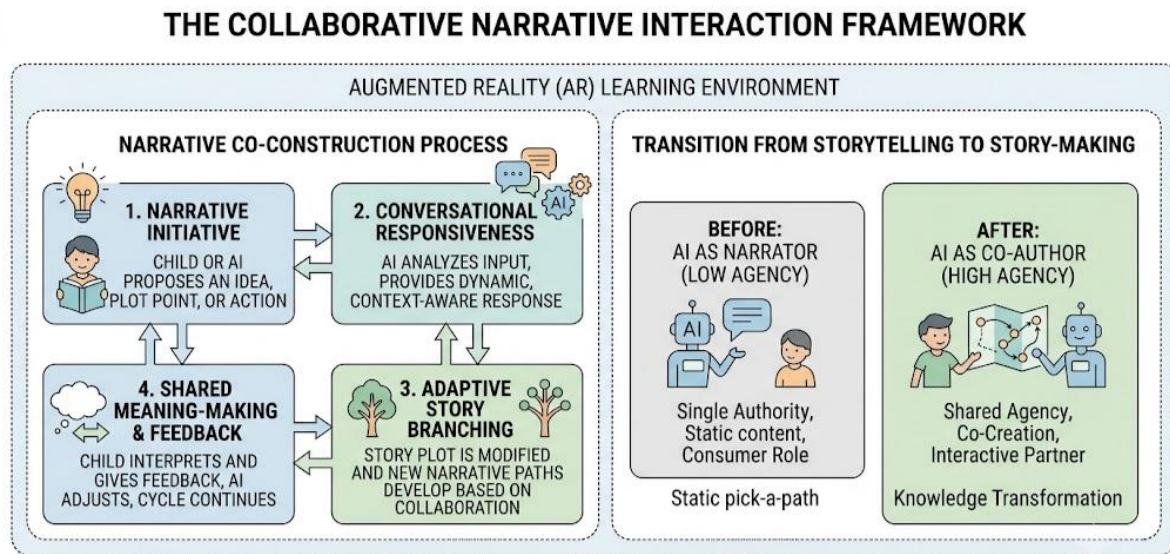


Figure 2. The Collaborative Narrative Interaction Framework

Narrative Initiative

Narrative initiative refers to the capacity, technological affordance, and pedagogical permission for both the child and the AI to dynamically initiate, create, shape, and abruptly redirect the story's trajectory through ongoing interaction. In traditional conversational agents, initiative is heavily skewed; the

system typically asks a closed question and waits for a specific categorical answer. In a co-constructive paradigm, initiative must be mixed, fluid, and reciprocal.

To facilitate narrative initiative, the ConvAI must utilise strategically open-ended prompts, moving beyond interrogative assessment to creative invitation. For example, rather than asking, "What colour is the forest?", the AI might prompt, "The forest is very dark and making strange sounds; what should our character use to find their way?". This requires the AI's NLP engine to comprehend highly imaginative, non-standard syntax and seamlessly weave those ideas into the evolving logic of the story.

Crucially, the AR interface allows the child to manifest their initiative physically and visually. If a child states, "I want a giant blue bird to carry us away," the system must not only acknowledge this conversationally but dynamically trigger the rendering of the corresponding AR asset in the physical space. However, maximising learner agency does not equate to total unconstrained freedom, which can lead to narrative collapse or cognitive exhaustion. The AI must exert a subtle structuring influence, acting as a narrative anchor, gently guiding the child back to a coherent plotline or pedagogical objective if the story becomes too fragmented, thereby perfectly balancing child-centric creativity with underlying instructional goals.

Conversational Responsiveness

Conversational responsiveness is defined as the technical and relational ability of the AI to actively listen, accurately interpret, and dynamically adapt to the child's input in a manner that feels authentic, empathetic, and contextually appropriate, maintaining a balanced dialogue. This component bridges the profound gap between basic speech-to-text transcription and deep, relational conversational understanding.

The foremost challenge in conversational responsiveness is mastering the acoustic and affective dynamics of children's speech. Early childhood speech is characterised by high pitch variability, phonological discrepancies, mispronunciations, short utterances, and rapid emotional volatility. Standard automatic speech recognition engines routinely fail under these conditions. The framework mandates the deployment of child-specific acoustic models, such as OpenAI's Whisper adapted for paediatric voices, paired with NLP engines trained to infer intent from the narrative context when direct phonetic transcription is unreliable (Fayz et al., 2025). Furthermore, incorporating emotion-recognition architectures enables the AI to detect user frustration, boredom, or excitement, allowing the agent to dynamically adjust its tone, conversational pacing, and level of scaffolding accordingly.

Beyond acoustic accuracy, the ConvAI must exhibit empathetic and relational attunement. Related work on relational AI, including the Significant Other AI (SO-AI) concept (Park, 2025), and chatbot-mediated mutual reminiscence, such as the Remini framework (Jiang et al., 2025), provides relevant design insights for emotionally responsive and reciprocal conversational interaction. When a child expresses a fear or a personal anecdote during the story-making process, the AI should respond with empathy, validating the child's emotion before integrating it into the narrative. This creates an emotionally supportive environment that fosters reciprocal self-disclosure and strengthens the child-agent collaborative bond. Additionally, maintaining conversational fluency requires minimising processing latency. The framework advocates for edge computing to handle basic interactions, paired with non-verbal visual cues, such as an AR character pondering, nodding, or gesturing, to mask processing delays and sustain the psychological illusion of continuous, unbroken turn-taking.

Adaptive Story Branching

Adaptive story branching is the structural mechanism by which the narrative logically, textually, and visually evolves in real-time based on the collaborative inputs of the child and the AI. Unlike static branching narratives commonly found in digital media, adaptive branching in this framework is generative, open-ended, and pedagogically driven.

This requires real-time contextual fusion. As the story diverges based on user input, both the generated dialogue and the spatial AR environment must remain perfectly synchronised. If the child decides the story should move from a forest to underwater, the ConvAI must instantly update the narrative context, and the AR engine must simultaneously alter the spatial visualisations (e.g., spawning aquatic flora, adjusting lighting algorithms, and triggering spatial bubble audio). This demands a centralised scene-state manager that coordinates shared data streams between the generative LLM and the spatial game engine (such as Unity3D).

Crucially, adaptive branching is not purely a mechanism for entertainment; it is deeply tied to the IDC habit loop and personalised, adaptive learning. The system continuously monitors the child's vocabulary usage, reading comprehension speed, and behavioural engagement levels. If a child demonstrates mastery of current linguistic concepts, the AI seamlessly branches the story into more complex socio-emotional themes or introduces advanced, discipline-specific vocabulary. Conversely, if the child hesitates or exhibits frustration, the AI branches into a supportive, explanatory loop, breaking down complex ideas. To maintain continuity over multiple sessions, the system employs narrative path-tracing. This allows the AI to recall past story-making sessions and reference previous co-constructed events, thereby rewarding the child's long-term engagement and validating their creative contributions as permanent fixtures of the persistent story world.

Illustrated AR-Based Learning Scenarios

To demonstrate the practical application of the CNIF, the following theoretical scenarios illustrate how ConvAI and AR function symbiotically as collaborative partners rather than narrative authorities.

Scenario A: Environmental Science and Vocabulary Acquisition

Context: A seven-year-old learner is exploring a story-making module centred on forest ecosystems.

Interaction:

- AR Visualisation: The physical living room is augmented with a sparse 3D forest floor. A small, animated AR fox appears on a rug.
- ConvAI (Initiative): "The little fox is very hungry, but winter is coming. How should we help her prepare?"
- Child (Co-Construction): "She needs to find food and hide it under the ground."
- ConvAI (Responsiveness & Scaffolding): "That's a brilliant idea! Finding and gathering food in the wild is called *foraging*. Let's help her forage. What kind of food should she look for?"
- Child (Initiative): "Berries and shiny beetles!"
- AR Visualisation (Contextual Fusion): The Unity3D engine dynamically spawns interactive AR berry bushes and glowing beetles across the room's physical surfaces.
- ConvAI (Adaptive Branching): "Look, there are berries on the couch! If she eats the beetles, she gets protein. If she eats the berries, she gets vitamins. Which path should our story take first?"

Analysis: In this scenario, the AI operates as a peer, introducing a vocabulary word (foraging) naturally within the child's established context. The visual AR environment responds instantaneously to the child's creative input, lowering the cognitive barrier to understanding abstract ecological concepts while fulfilling the creation loop of the IDC theory.

Scenario B: Social-Emotional Learning and Empathy Building

Context: A five-year-old child is engaging with a fantasy narrative focusing on interpersonal conflict resolution.

Interaction:

- AR Visualisation: An animated AR robot is sitting on the floor, its visual display flashing blue, accompanied by a low, melancholic audio tone.
- ConvAI (Relational Attunement): "Our friend Sparky looks very sad today. He tried to build a tower, but the other robots knocked it down. What should we say to him?"

- Child (Empathy Generation): "It's okay Sparky, I will help you build a bigger one so they can't break it."
- ConvAI (Responsiveness): "That is very kind of you. Sparky feels better knowing he has a friend to help. But he is still a bit angry at the others. Should we teach him to take a deep breath, or should we go talk to the other robots?"
- Child (Decision Making): "Let's take a deep breath first."
- AR Visualisation (Contextual Fusion): The AR robot stands up, mimicking a deep breathing animation, and its display shifts from blue to a calm green.

Analysis: In this context, the AI leverages affective narrative design to prompt the child into a state of empathy. The framework utilises narrative co-construction not just for literacy, but also for socio-emotional scaffolding. By providing the child with shared agency over the emotional resolution, the system offers practical, experiential practice in emotional regulation and problem-solving without the high stakes of real-world social conflict.

Transformational Impacts on Learning and Development

Transitioning to a story-making paradigm through the CNIF yields multi-dimensional educational benefits that extend far beyond foundational literacy. By positioning the child as an active, agentic co-author in a multimodal learning ecosystem, the framework facilitates complex cognitive development, executive functioning, and identity formation.

Advancing Literacy and Linguistic Comprehension

The primary objective of the framework is to enhance emergent literacy and language skills. Through continuous, dynamic dialogue with the ConvAI, children practise expressive language, rapid vocabulary retrieval, and complex syntactic structuring in a highly contextualised, meaningful environment. When a child encounters an unfamiliar concept or word in the story, the AI can immediately provide both an AR visual representation and a conversational, age-appropriate definition, seamlessly bridging the gap between abstract text and concrete, spatial understanding.

Moreover, the act of narrative co-construction inherently demands that the child understand and manipulate narrative sequencing (beginning, middle, end), cause-and-effect relationships, and character motivations. By actively deciding what happens next and facing the logical consequences of those decisions within the story world, the child engages in higher-order reading comprehension. They transition from merely decoding written words to synthesising deeper narrative meaning.

Enhancing Executive Function and Cognitive Flexibility

Conversational exchanges during story-making are uniquely positioned to enhance EF. Working memory is rigorously exercised as the child must retain the context of the story, remember previous plot points, and hold character traits in mind while generating new ideas. Inhibitory control is practised through the structured turn-taking required to interact effectively with the AI, teaching the child patience and the rhythms of conversational exchange.

Cognitive flexibility is constantly demanded, as the child must adapt their thought processes when the AI introduces unexpected plot twists, asks challenging questions, or introduces a new visual element into the AR space. The AI's ability to provide adaptive scaffolding, adjusting its level of support based on real-time computational assessments of the child's cognitive load, helps to keep the learner within their zone of proximal development, fostering resilient cognitive growth.

Fostering Motivation, Self-Efficacy and Identity

Under traditional storytelling models, learners who struggle with reading frequently develop avoidance behaviors, anxiety, and low academic self-efficacy due to repeated experiences of frustration. Story-making fundamentally alters this dynamic by shifting the pedagogical focus from "getting the right answer" to "creating something novel and valuable".

The highly personalised nature of the generated narratives, dynamically tailored to the child's specific interests (e.g., incorporating their favourite animals, hobbies, or cultural touchstones), dramatically amplifies intrinsic motivation. As the child witnesses their spoken ideas instantly manifested as 3D AR objects and seamlessly integrated into the AI's storyline, they experience a profound sense of agency, competence, and accomplishment. This process of creative expression allows children to explore their inner world, articulate their developing values, and safely navigate complex emotions. As media creation becomes more accessible, children can act as "cyberflâneurs" (Marsh, 2016), navigating and shaping digital spaces, which contributes significantly to their overarching identity development and psychological resilience.

Future Directions and Cross-Disciplinary Applications

While the CNIF was conceived primarily within the context of early childhood literacy, its underlying principles of shared agency, multimodal visualisation, and conversational scaffolding are highly extensible. The framework provides a versatile foundation for cross-disciplinary applications across the educational spectrum.

Science, Technology, Engineering, and Mathematics (STEM) Education and Collaborative Scientific Inquiry

In Science, Technology, Engineering, and Mathematics (STEM) domains, the CNIF can transform how students interact with abstract, imperceptible phenomena. Instead of co-constructing a fictional narrative, the learner and the agentic AI can co-construct a scientific simulation. For example, a student exploring physics could instruct the AI to alter the gravity of a virtual AR environment. The AR interface would visually render the resulting kinetic changes, while the ConvAI asks hypothesis-driven questions (e.g., "What do you think will happen to the trajectory of the ball if we double the mass?"). This active, conversational inquiry models the scientific method, transforming passive observation into collaborative, rigorous experimentation, aligning perfectly with the ICAP framework's highest tiers of cognitive engagement.

Language Acquisition and Bilingual Education

Language learning and bilingual education can benefit immensely from the framework's capabilities. AR can create immersive, contextual environments, such as a virtual marketplace or a historical cultural setting, for practising conversational exchanges. The ConvAI serves as an untiring, non-judgemental conversational partner, offering real-time feedback on pronunciation, grammar, and vocabulary usage within the context of an engaging story. By leveraging IDC theory, the system ensures language activities are deeply relevant to learners' interests, encouraging the consistent, low-anxiety practice necessary for true fluency.

Limitations of the Proposed Framework

While the CNIF offers a robust conceptual architecture for next-generation learning environments, several limitations warrant explicit acknowledgement. Outlining these constraints provides necessary boundaries for the framework's application and clear directions for subsequent empirical evaluation.

First, as this is a conceptual article grounded in theory synthesis, its arguments are derived primarily from a curated selection of literature spanning computer-supported collaborative learning, cognitive psychology, and human-computer interaction. Since generative AI is evolving at an unprecedented pace, literature regarding agentic AI co-authorship remains emergent and highly fluid.

Second, the CNIF currently exists as a theoretical blueprint. While the pedagogical foundations, such as the ICAP framework and IDC theory, have been empirically validated in isolation, their integrated application within a simultaneous AR/ConvAI environment requires rigorous field testing. There is a need for longitudinal, controlled empirical studies to determine whether the posited improvements in executive functioning, vocabulary acquisition, and reading comprehension occur

consistently across diverse demographic groups. Future research should prioritise mixed-methods evaluations that combine quantitative learning analytics with qualitative observational data to assess the efficacy of adaptive story branching in real-world classroom settings.

Finally, the technical feasibility and interaction quality of the framework may vary across children's ages, languages, accents, developmental profiles, devices, and contexts. Child speech, for example, remains challenging for general-purpose automatic speech recognition, and performance can depend substantially on the characteristics of the child-speech data used for evaluation or adaptation. Future research should therefore operationalise each CNIF component, evaluate the framework with diverse child populations, compare co-authoring interactions with conventional narrator or predefined branching conditions, and assess learning outcomes, interaction quality, safety, usability, and longer-term developmental effects empirically.

The Imperative for Longitudinal Empirical Research

Current evaluations of AI and AR in education predominantly consist of short-term usability studies, localised pilot programmes, or theoretical postulations. To rigorously validate the efficacy of the CNIF, large-scale, longitudinal empirical studies are an urgent necessity. Researchers must measure not only immediate engagement metrics and short-term vocabulary acquisition but also long-term, sustained impacts on EF, critical thinking, reading comprehension, and the development of lifelong learning habits. Furthermore, analysing the longitudinal psychological and sociological effects of human-AI collaboration will be crucial for the continuous refinement of ethical guidelines surrounding the deployment of increasingly autonomous digital companions in the lives of children.

Conclusion

The integration of AR and ConvAI represents a watershed moment in the evolution of educational technology, perfectly encapsulating the imperative to reimagine creativity through AI. However, realising the profound potential of these transformative tools requires moving beyond the mere digitisation of traditional, didactic instructional methods. By conceptualising AI not as a static narrator, an automated evaluator, or a content delivery mechanism, but as a reciprocal, agentic partner in meaning-making, the educational landscape can be fundamentally reshaped.

The CNIF introduced in this paper provides a comprehensive, theoretically grounded blueprint for actualising this paradigm shift. By operationalising narrative initiative, conversational responsiveness, and adaptive story branching, the framework is designed to elevate children from passive consumers to active, empowered co-authors of their educational journeys. Grounded in robust pedagogical theories such as the IDC theory, the CTML, and the ICAP framework, this approach

meticulously balances imaginative freedom with essential cognitive scaffolding, aiming to translate high engagement into measurable developmental gains, a link that future empirical work must test.

Overcoming the inherent technical challenges of multi-modal synchronisation, child-specific speech recognition, and mobile hardware optimisation requires rigorous, ongoing interdisciplinary collaboration between computer scientists, interaction designers, and educational psychologists. Equally paramount is an uncompromising commitment to ethical, child-centred design, safeguarding data privacy, actively mitigating algorithmic bias, and maintaining healthy relational boundaries between the developing child and the machine.

Ultimately, the shift from storytelling to story-making democratises the creative process for early learners. It cultivates their linguistic abilities, EF, and digital literacy, while simultaneously validating their creative agency and identity. By designing AI systems that genuinely listen, intelligently adapt, and collaboratively create, we can build immersive educational environments that not only teach children how to read the world, but actively provide them with the collaborative tools to reimagine and rewrite it.

References

- [1] Aligned. (n.d.). A brief history of literacy instruction in America. <https://wearealigned.org/brief-history-literacy-instruction-america/>.
- [2] Azhar, M., Hew, S. H., & Neo, T.-K. (2024, December). SPARC: A framework for enhancing early childhood reading through augmented reality, conversational AI, and IDC theory. In 2024 International Conference on TVET Excellence & Development (ICTeD) (pp. 103-106). IEEE.
- [3] Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4), 355–385.
- [4] Bandura, A. (1977). *Social learning theory*. Prentice Hall.
- [5] Beneteau, E., Richards, O. K., Zhang, M., Kientz, J. A., Yip, J., & Hiniker, A. (2019). Communication breakdowns between families and Alexa. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), Article 243.
- [6] Chan, T. W., Looi, C. K., Chang, B., Chen, W., Wong, L. H., Wong, S. L., ... & Chou, C. Y. (2019). IDC theory: creation and the creation loop. *Research and Practice in Technology Enhanced Learning*, 14(1), 26.
- [7] Chan, T. W., Looi, C. K., Chen, W., Wong, L. H., Chang, B., Liao, C. C., ... & Ogata, H. (2018). Interest-driven creator theory: Towards a theory of learning design for Asia in the twenty-first century. *Journal of Computers in Education*, 5(4), 435-461.

- [8] Chen, W., Chan, T. W., Wong, L. H., Looi, C. K., Liao, C. C., Cheng, H. N., ... & Pi, Z. (2020). IDC theory: habit and the habit loop. *Research and Practice in Technology Enhanced Learning*, 15(1), 10.
- [9] Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243.
- [10] Copple, C., & Bredekamp, S. (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8* (3rd ed.). National Association for the Education of Young Children.
- [11] Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning*, 18(4), 607-614.
- [12] Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, 210, 104967.
- [13] Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64(1), 135-168.
- [14] Fayz, S., Haita, A., Sabiri, M., & Qaraai, Y. (2025, April). Fine-Tuning Whisper for Children's Speech Recognition: Enhancing ASR Accuracy with Specialized Data. In *International Conference on Innovative Approaches and Applications for Sustainable Development* (pp. 457-460). Cham: Springer Nature Switzerland.
- [15] Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.
- [16] Jaakkola, E. (2020). Designing conceptual articles: four approaches. *AMS Review*, 10(1), 18-26.
- [17] Jiang, Z., Yeo, S., Seow, W. X. D., & Perrault, S. T. (2025). Remini: Leveraging Chatbot-Mediated Mutual Reminiscence for Promoting Positive Affect and Feeling of Connectedness among Loved Ones. *Proceedings of the ACM on Human-Computer Interaction*, 9(7), 1–43.
- [18] Lerkkanen, M.-K., Kiuru, N., Pakarinen, E., Poikkeus, A.-M., Rasku-Puttonen, H., Siekkinen, M., & Nurmi, J.-E. (2016). Child-centered versus teacher-directed teaching practices: Associations with the development of academic skills in the first grade at school. *Early Childhood Research Quarterly*, 36, 145–156.
- [19] Liu, X., Xing, Y., Chen, Y., Pei, W., Choi, J., & Xu, Y. (2025). Children's Narrative Participation in Story Co-Creation with AI vs. Human Partners: Response Styles and Narrative Features. In *Proceedings of the 24th Interaction Design and Children* (pp. 803-809).
- [20] Lovato, S. B., & Piper, A. M. (2015). "Siri, is this you?": Understanding young children's interactions with voice input systems. In *Proceedings of the 14th International Conference on Interaction Design and Children* (pp. 335–338). Association for Computing Machinery.
- [21] Mai, X. (2025). The peer in the machine: Evaluating the impact of a "peer-like" AI persona on writing motivation in low-proficiency L2 learners. *Open Journal of Applied Sciences*, 15(11), 3624–3635.

- [22] Marcon, R. A. (2002). Moving up the grades: Relationship between preschool model and later school success. *Early Childhood Research & Practice*, 4(1).
- [23] Marsh, J. (2016). ‘Unboxing’ videos: co-construction of the child as cyberflâneur. *Discourse: Studies in the Cultural Politics of Education*, 37(3), 369-380.
- [24] Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.
- [25] McTear, M. (2022). *Conversational AI: Dialogue systems, conversational agents, and chatbots*. Springer.
- [26] Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, 137(2), 267–296.
- [27] Monjoree, U., Aguilar, S. D., Qian, C., Van Huyck, C., Yeh, S. H., Tranbarger, P., ... & Yan, W. (2024). AR-Classroom: Integrating conversational artificial intelligence with augmented reality technology for learning spatial transformations and their matrix representation. 2024 IEEE Frontiers in Education Conference (FIE), 1–8.
- [28] Mukherjee, A., & Chang, H. H. (2025). Agentic AI: Autonomy, accountability, and the algorithmic society. arXiv preprint arXiv:2502.00289.
- [29] Ojeda-Ramirez, S., Han, A., Rich, A., Han, S., & Peppler, K. (2025). Reflective AI-Partnerships: How Middle Schoolers Balance Creativity and AI Collaboration. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025*, pp. 404-408. International Society of the Learning Sciences.
- [30] Park, S. (2025). Significant Other AI: Identity, Memory, and Emotional Regulation as Long-Term Relational Intelligence. arXiv preprint arXiv:2512.00418.
- [31] Robertson, J., & Good, J. (2005). Story creation in virtual game worlds. *Communications of the ACM*, 48(1), 61-65.
- [32] Stahl, G., Koschmann, T. D., & Suthers, D. D. (2006). CSCL: An historical perspective. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge University Press.
- [33] Stipek, D. (2006). No Child Left Behind comes to preschool. *The Elementary School Journal*, 106(5), 455–465.
- [34] Strickland, D. S., & Riley-Ayers, S. (2006). Early literacy: Policy and practice in the preschool years. *National Institute for Early Education Research*.
- [35] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.
- [36] Wong, L. H., Chan, T. W., Chen, W., Looi, C. K., Chen, Z. H., Liao, C. C., ... & Wong, S. L. (2020). IDC theory: interest and the interest loop. *Research and Practice in Technology Enhanced Learning*, 15(1), 3.

- [37] Wu, Y., Mahmud, R., Bin Sawad, A., van Den Hoven, E., & Kocaballi, A. B. (2026). Conversational agents in augmented reality: A systematic review. *ACM Transactions on Interactive Intelligent Systems*.
- [38] Zhang, Z., Xu, Y., Wang, Y., Yao, B., Ritchie, D., Wu, T., ... & Li, T. J. J. (2022, April). StoryBuddy: A human-AI collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-21).

Acknowledgment

The author would like to express sincere gratitude to all individuals and institutions who have directly or indirectly contributed to the completion of this research and the preparation of this manuscript. Their support, guidance, encouragement, and valuable contributions have been instrumental in the successful conduct of this study. The author deeply appreciates the assistance and cooperation received throughout the research process.

Funding Information

This research is supported by the Ministry of Higher Education Fundamental Research Grant Scheme (FRGS) (Grant No. FRGS/1/2022/SSI0/MMU/01/1). The authors would also like to thank the students and staff of the Faculty of Creative Multimedia, Multimedia University, for supporting this research study. The authors would also like to express their appreciation to the TM R&D Grant fund from TM R&D Sdn. Bhd., MMU (MMUE/230053 / RDTC/231106) for their support and funding, and thank all the members for their effort, contribution, and collaboration in this study.

AI and LLM Disclosure (Limited Use)

Generative AI and LLM tools were utilized solely to support language enhancement, editing, and manuscript formatting. These tools were not used to generate research findings, analyses, interpretations, or conclusions. All AI-assisted text underwent substantial human review and revision, and the final manuscript was critically evaluated and approved by the authors. The authors retain full responsibility for the integrity, originality, accuracy, and scholarly contribution of this work.

Authors' Bio

Masood Azhar is a Master's by Research candidate at the Faculty of Creative Multimedia, MMU, Malaysia. His research interests include AI, LLMs, immersive technologies, educational technology, personalized learning, human-computer interaction, and digital learning environments. His current

research explores the integration of emerging technologies to enhance learner engagement, motivation, and educational experiences.

Assoc. Prof. Dr. Hew Soon Hin is an Associate Professor at the Faculty of Creative Multimedia, MMU, Malaysia. His research interests include computer-assisted instruction, computer-assisted language learning, instructional design, educational technology, interactive multimedia, storytelling, and digital learning environments. He has published extensively in the areas of technology-enhanced learning and multimedia-based educational applications. His research focuses on leveraging interactive technologies to improve learning experiences, language acquisition, reading engagement, and educational outcomes. He has also led and contributed to the development of various educational applications and digital learning solutions for language and multimedia education.

Prof. Ts. Dr. Neo Tse Kian is a Professor at the Faculty of Creative Multimedia, MMU, Malaysia, and a member of CAMELOT. His research specializes in multimedia technology in education, constructivist learning environments, educational technology, interactive multimedia, and learner-centered pedagogical approaches. He has extensive experience in academic leadership, research supervision, curriculum development, and educational innovation. His work focuses on the design and implementation of technology-enhanced learning environments that promote active learning, collaboration, creativity, and meaningful learner engagement through multimedia technologies.