An Automated Content Generation Framework for Extended Reality Based Safety Training in Construction

Yulun Wu[1], Gaang Lee[1], Qipei Mei[1], Claudio Mourgues[3], Vicente A. Gonzalez[2]

[1]Infrastructure and Human Tech (IHT) Lab, Dept. of Civil and Environmental Engineering, Faculty of Engineering, University of Alberta, Edmonton, Alberta, Canada
[2]Infrastructure and Human Tech (IHT) Lab, Strategic Projects Insight Centre in Engineering (SPICE), Dept. of Civil and Environmental Engineering, Faculty of Engineering, University of Alberta, Edmonton, Alberta, Canada
[3]Dept. of Construction Engineering and Management, Pontificia Universidad Catolica de Chile, Santiago, Chile
yulun6@ualberta.ca, gaang@ualberta.ca, qipei@ualberta.ca, cmourgue@uc.cl, vagonzal@ualberta.ca

Abstract

The construction industry has one of the highest rates of workplace fatalities and severe injuries. Effective, high-fidelity safety training is not merely a regulatory requirement but a moral imperative. While Extended Reality (XR) has proven to improve the training results in construction safety training, a bottleneck is its resource-intensive content creation processes that require multidisciplinary collaboration among subject matter experts, pedagogical designers, 3D modelers, and software developers, resulting in development timelines spanning months and prohibitive costs. Authoring tools have been used to solve this issue, allowing development in a no-code or low-code environment, while they remain constrained by templated rigidity, asset dependencies, lack of adaptability and automated regulatory compliance integration. This paper addresses these gaps by proposing a novel framework for automated XR authoring that integrates Large Language models (LLMs) and Procedural Content Generation (PCG). The framework transforms natural language into interactive, rule-compliant training instances through four core contributions. (1) An LLM-to-machine executable pipeline that translates unstructured instruction into structured, machine-interpretable instructions. (2) A PCG engine that automatically arranges 3D assets and environmental hazards based on site-specific logic and safety constraints. (3) A DKT-driven adaptive engine that personalize training module by dynamically adjusting scenario difficulty. (4) A multimodal evaluation method that combining subjective and objective training data for validating the usability and pedagogical effectiveness of the generated content. This paper outlines the context, framework architecture, procedural generation pipelines, and validation strategies required for creating and implementing the authoring tool, towards the next generation of XR based construction safety training.

Keywords –

Construction Safety Training; Authoring Tool; Extended Reality; Large Language Models; Artificial Intelligence; Procedural Content Generation

1 Introduction

The construction industry has one of the highest rates of workplace fatalities and severe injuries [1, 2]. The "Fatal Four" hazards, including falls, struck-by incidents, caught-in and caught-between accidents, and electrocutions, categorize the main areas of workforce hazard recognition safety training [3]. While classroom training dominate, Extended Reality (XR) based training is increasingly adopted for it high-fidelity immersion without real-world risks [4]. However, many organizations, particularly small-to-medium enterprises or schools, do not have the expertise or financial budget required and avoid investing in XR devices and training programs [5]. Besides, creating a site-specific safety training module requires collaborative work by Subject Matter Expert (SME) who can define exact hazards and procedures to be trained, the pedagogy experts who develop the scenarios, feedback mechanisms and assessment methods, and technical expertise in 3D modelling, gaming development, user experience and user interface designs [6, 7]. Authoring tools have been developed to allow SMEs developing XR based training content within a no-code or low-code environment with the possibility of modification [8, 9]. After the initial development, the resulting training modules are often static, offering minimal flexibility for unique training preferences of individual trainees [10]. This study proposes an automated authoring framework that further streamlines the process of content creation of a construction safety training module. Several gaps are identified in this study. (1) Lack of automated content generation. Current authoring tools require manual template selection, asset replacement, scenario configuration with limited scalability and requiring techical expertise. (2) Absense of automated regulatory compliance. Existing authoring systems are not capable to validate training modules against specific safety regulations, compliance checking remains a manual process requiring expert review. (3) Limited adaptive learning capabilities. Once developed, XR training module are predominatly static, offering minimal personaliztion based on individual learner performance or learning pace. (4) Insufficient integration of multimodal assessment. Existing XR module rely on post-training knowledge tests, with less focusing on integrating conginitve load, situational awareness or other physiological responses during training. The following objectives directly mapped to the identified gaps, including (1) Formalize an natural language processing pipeline that translate natural language into machine-intepretable instructions. (2) Develop a regulatory constraint logic that aligns generated content with domestic safety code or regulations. (3) Design an adaptive learning arfhitecture to personalize training difficulty dynamically. (4) Establish a multimodal evaluation method that integrates physiological metrics into the trainig assessment loop. By incorporating LLMs and PCG, the complexity of content creation and adaptability will be reduced further. This leads to our primary research question: How can LLM and PCG be formalized into an automated authoring tool that is capable of transforming natural language instructions into high-fidelity and rule-compliant XR safety training scenarios?

2 Related Works

The Related Works section provides conceptual context to four technological domains that are relevant to this research including the landscape and limitations of current authoring tools, state-of-the-art of LLMs on their ability to interpret safety regulations and natural language input, the usage of PCG to automate the creation of spatial environments, the human elements and pedagogy design that need to be considered.

2.1 Authoring Tools for XR Development

The development of XR training content has historically followed two distinct workflows. High-complexity pipelines rely on full-featured game engines such as Unity and Unreal Engine, which demand advanced programming and 3D modeling expertise. To address this, no-code or low-code authoring platforms such as Thinklink, Uptale (formerly VRBuildier), Zapworks have integrate drag-and-drop editor interfaces, built-in model libraries, 360-degree media footage, and provide dashboards to track trainees' performance [11-13]. Academic efforts have similarly produced domain-specific tools that allow narrative creation without coding [8, 14-16]. However, most solutions remain manual as the SME must still explicitly define every scene, place every model, and script every interaction, resulting in static scenarios that fail to reflect the dynamic nature of construction sites.

2.2 The Role of Large Language Models in Authoring

Contemporary authoring ecosystems increasingly incorporate LLMs. Platforms such as MotiveAI generate personalized dialogue and coaching feedback, while SkyboxAI produces 360-degree environments from text prompts [17, 18]. Inworld AI and Ubisoft's NEONPC projects integrate LLMs into game engines to drive Non-Player Characters (NPCs) with unscripted behaviour [19, 20]. Research prototypes further demonstrate LLM-driven text-to-scene generation (SceneCraft, LayoutGPT) that converts descriptions such as "a cluttered site with a fall hazard" into structured 3D layouts [21, 22]. Additional studies illustrate LLM generation of behaviour trees for NPCs and automated compliance checking of regulatory texts against Building Information Models (BIM) [23-25]. LLMs have also been integrated into VR platforms for processing construction site imagery to identify safety violations [26]. These solutions operate as black-box systems lacking a modular, transparent architecture tailored to construction safety training. The absense of standardized data pipeline between LLM semantic engines and XR rendering environment continues to hinder reproducebility and regulatory verifiability [27].

2.3 The Role of Procedural Content Generation in Authoring

Procedural Content Generation enables algorithmic creation of diverse training environments, ensuring no two sessions are identical, therefore supporting individualized skill transfer [28, 29]. Algorithm such as wave function collapse have proven effective for constructing rule-complaint assemblies (e.g., scaffolding or piping networks) [30]. When combined with reinforcement learning, PCG can dynamically adjust environmental complexity to match trainee proficiency, as demonstrated in fire-evacuation and building-collapse scenarios [31, 32]. In safety-critical applications, PCG must tightlky constrained by domain specific regulatory logic rather than purely aesthetic or entertainment purposes. Although LLMs excel at high-level intent interpretation, bridging the semantic gap between LLM outputs and the low-level geometric constraints of PCG engines remains a central technical challenge, leading to semantic gaps where the genrated 3D scenes fails to match the author's specific safety instructions [33].

2.4 Human Elements and Pedagogy Designs

Effective XR training is grounded in andragogical principles that treat adult learners as self-directed participants rather than passive recipients [34]. Situated Learning Theory underscores the necessity of embedding knowledge acquisition within authentic contextual practice [35], while Embodied Cognition theory explains how physical interaction with the virtual world reinforces hazard recognition skills [36]. Human-Centred Design further emphasize careful management of cognitive load so that trainee's mental resources are focused on high-risk decision-making rather than interface navigation [37]. Real-time physiological monitoring such as eye-tracking or heart rate variability enables a system to maintain the trainee in the "Zone of Proximal Development," preventing both boredom and cognitive overload [38, 39].

3 Framework Architecture

Recent advances in LLMs and PCG have demonstrated isolated capabilities in text-to-scene generation, rule interpretation, and procedural environment synthesis. The proposed framework is initiated by the challenges discussed and an XR design principles roadmap [40], which identifies SME natural language input as the input and foundation. Captured data integration follows recommendations for XR content generation [41]. Topic-specific templates address tmplate rigidity challenges and directly drawn from the existing autrhoing tools [42]. The data and asset management layer resolves asset dependency issue noted in Section 3. As illustrated in Figure 1, the architecture begins with SME natural language input as the foundation and incorporates captured data, topic-specific templates, and 3D asset libraries in accordance with established XR design principles. hese inputs flow through a data and asset management layer that resolves asset dependencies before entering the generative module. The framework further includes an adaptive feedback loop and evaluation components that enable iterative personalization while maintaining human oversight at every critical checkpoint.

Figure 1 includes front-end interfaces (blue blocks) for authoring and deployment, back-end processing engines (orange blocks) comprising the semantic engine for prompt interpretation, the spatial engine for procedural 3D environment generation, and the social engine for NPC behaviour and dialogue management, a regulatory constraint solver (gold hexagon) that enforces safety-code compliance on all outputs, and green components representing the adaptive engine and multimodal evaluation loop that feed trainee performance data back into the system for variation generation. Blocks marked with "H" denote mandatory human-in-the-loop intervention points where subject-matter experts can review and modify generated content before deployment. Runtime data are stored in the data and asset management layer to support subsequent personalized training cycles.

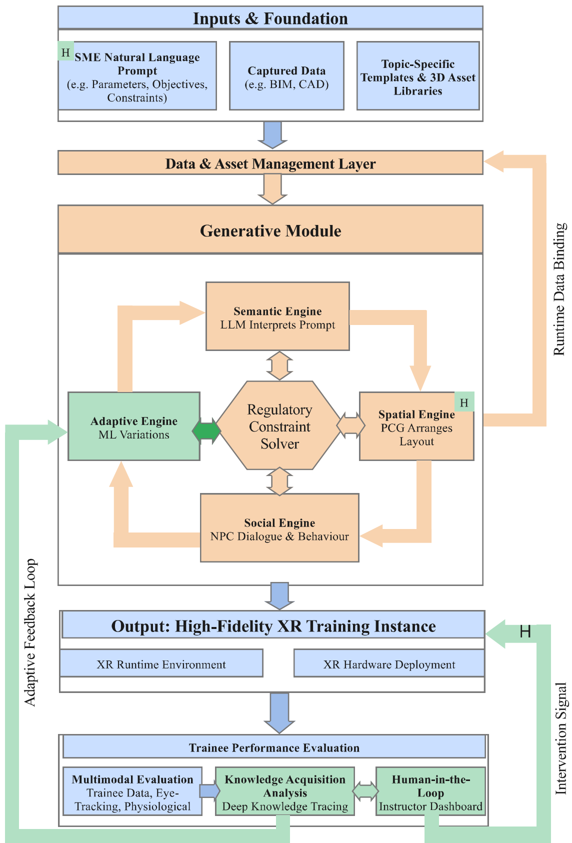

Figure 1. Automated Content Generation Framework for XR Safety Training

3.1 Architecture of the Authoring Tool

SME natural language input and topic-specific templates can reduce tehcnical expertise [40] and no-code cases [8], captured data and asset management directly address content generation dependencies that identified by researchers [41]. An developed automated authoring tool relies on pre-validated, topic-specific templates that correspond to specific safety modules (e.g., excavation safety, steel erection, electrical lockout). When a training topic is selected, the system instantiates a foundational environment that allows the generative module to focus on variable hazards rather than rebuilding the static geometry. This semantic grounding requires initial SME validation, enables the system to instantiate a knowledge-aware foundational environment, consequently bypass the high-latency task of rebuilding static geometry.

To accommodate organizations of varying sizes, the framework allows system supporting automatic retrieval of 3D models from open-source libraries for smaller entities while also accepting direct ingestion of captured data, such as LiDAR scans and BIM/CAD imports, for organizations with established asset repositories. Realism levels are calibratted according to specific training objectives, basic hazard recognition may require abstract representations, complex equipment operations will require more detailed assets in terms of complexity of interaction and rendering quality. Once a trainee completes at least one training cycle, an encrypted performance profile is created within the data and asset management layer, which functions as the central synchronization hub. This profile acts as a seed for the adaptive engines, so that subsequent scenarios can be tailored to the individual learner's competency level. For a novice, the system injects high-contrast hazard assets and visible cues to lower the difficulty level, into the spatial engine's layout logic. For an expert, the seed triggers the social engine to generate latent hazards, such as subtle environmental cues and unscripted NPC behaviours to test advanced situational awareness. All performance metrics should be processed locally with encryption to preserve data privacy and are made available for SME intervention when pedagogical calibration is required.

3.2 The Generative Module for Training Instances

The generation of each specific training instance is driven by synergistic loops among the semantic engine [21, 22], the spatial engine [28], the social engine [24] and the adaptive engine [10], which were designed to overcome the manual workflow, static scenario development, and limited regulatory compliance capabilities identified earlier.

The semantic engine begins with the SME providing a natural language prompt, which must include training objective, difficulty level, and environmental constraints. For example, "Create a scaffold inspection scenario for a novice worker with high wind conditions." Drawing on the success applications such as SceneCraft [21] and LayoutGPT [22], the engine translates unstructured textual descriptions into structured, machine-interpretable parameters. These parameters are then passed to the spatial engine, which applies PCG to retrieve the appropriate template and procedurally arranges the 3D objects while enforcing built-in constraints derived from local safety norms and standards. For example, a generated guardrail needs to meet height requirements unless the training objective is to identify a defect. Such techniques have been successfully implemented in serious games and emergency training environments [28, 31].

The social engine incorporates LLM-driven NPC avatars that generate real time behavior trees and dialogue in real-time, creating the training environment for adaptive role-play. For example, an NPC that incorporates retrieval-augmented generation (RAG) can access a database of safety regulations to explain complex procedures in simple terms.[25] Adaptive Engine employs machine learning algorithms to introduce such as changes in weather, the type of hazard, or tool replacement, ensuring each session remains unique and aligned with trainee's evlving performance profile.

Every generated instances are routed through the regulatory constraint solver before deployment, guaranteeing that the simulation remains both visually plausible and legally compliant unless the scenario is intentionally constructed to test hazard-identification skills.

3.3 Multimodal Evaluation on Trainee Performance

Trainee proficiency across repeated sessions is continuously modeled by Deep Knowledge Tracing (DKT) algorithm, which is formulated as a recurrent neural network that utilizes long short-term memory units to capture long-range dependencies in trainee learning patterns [43]. Each trainee's interaction at time t as a tuple $x_t = \{q_t, a_t\}$, where q_t is the specific hazard type and a_t is the binary performance outcome, success or failure. The trainee's latent knowledge state vector h_t is updated via the transition function as below,

$$h_t = \tanh\,(W_{xh}x_t + W_{hh}h_{(t-1)} + b_h),$$

where W_{xh} and W_{hh} are the weight matrices that determine how much the current interaction x_t and the previous knowledge state h_{t-1} influence the new state, where b_h is the bias for the hidden layer. The probability of correct identififcation of future hazard is predicted by

$$y_t = \sigma\big(W_{hy}h_t + b_y\big),$$

such that y_t is the intervention signal that is a value between 0 and 1. A low predicted proficiency value suggests the trainee is struggling with a specific concept, serves as the intervention signal that adjusts the spatial and social engine parameters in the next generated training instance, modifing the density of hazards or visual saliency as required.

In parallel, the framework suggests recording physiological and behavioral data such as time-to-fixation, dwell time, pupil diameter, and heart-rate variability to distinguish genuine hazard recognition from superficial glances and to quantify cognitive load in real time [44]. These multimodal data streams feed directly into the adaptive feedback loop, ensuring the intervention signal is based on authentic physiological recognition rather than randomness. Furthermore, evaluation data is suggested to be synthesized into visual evaluation dashboards that highlight trainees' performance, so that trainees receive immediate, gamified feedback to foster engagement and cognitive alignment [45]. Trainers can also use this information for reviewing and evaluation purposes. The cognitive load measurements ensures the trainee is not being pushed into cognitive overload that pauses learning [46]. When DKT suggest increasing hazard complexity based on a high probability of success the cognitive load assessment may deliver constraint to the regulatory constraint solver, to keep the trainee in the "zone of proximal development" [47]. If the biosensors detect a spike in tonic pupil dilation beyond a pre-validated baseline, the adaptive feedback loop triggers a complexity ceiling. This prevents the generative module from introducing further extraneous load.

The collection and storage of individuals' data should be governed by a strict privacy-preserving architecture [40]. The entire data collection and storage process adheres to a strict privacy-preserving architecture that employs local differential privacy and edge-based processing, thereby protecting individual records against reverse-engineering while still allowing aggregated improvements to the training models. A mandatory human-in-the-loop protocol further guarantees that instructors retain final authority to review AI-generated scenarios and override assessment outcomes, ensuring alignment with organizational safety culture.

3.4 Validation Strategies for The Developed Authoring Tool

Validation of the automated authoring tool employs a multi-layered protocol that simultaneously assesses technical performance, regulatory compliance, usability, and pedagogical effectiveness. This section provide suggestions that researchers and practitioners should consider depends on their use cases.

Technical validation assesses system performance and content generation capabilities using established XR benchmarking metrics. Key performance indicators include the following, (1) content generation time, for example, with target reduction from weeks (traditional manual authoring) to under 2 hours per scenario, consistent with documented 5× time reductions achieved by recent XR authoring systems [48]; (2) system stability measured through crash-free operation across 50+ generation cycles, addressing latency and tracking accuracy requirements identified in XR device benchmarking studies [49]; and (3) asset retrieval accuracy targeting >85% semantic match between natural language queries and retrieved 3D models. Performance validation will employ hardware configurations meeting established XR content generation requirements (GPU ≥16GB VRAM, 32GB RAM) and maintain frame rates of 72–90 Hz with motion-to-photon latency <20ms as recommended for standalone XR headsets [50].

Regulatory compliance validation employs a single SME to verify that generated scenarios meet the creator's intended safety training objectives and regulatory requirements. This user-centric validation approach recognizes that the SME who authors the natural language scenario description possesses the most accurate understanding of their specific training needs and applicable regulatory context [51, 52]. The SME reviewer needs to have recognized safety certification or equivalent (i.e. ≥5 years construction safety experience [53]) evaluates generated scenarios against their original intent and specific safety codes.

As illustrated in the framework architecture (Figure 1), human validation occurs at strategic checkpoints marked with "H". The SME intervenes after automated LLM-based scenario generation and PCG asset placement but before deployment to trainees. This single-checkpoint validation replaces the current practice of continuous multidisciplinary team involvement throughout the entire authoring process. By concentrating human expertise at a single validation checkpoint rather than throughout the entire development cycle, the framework could achieve substantial time reduction, the SME reviews a nearly-complete scenario in 30-60 minutes rather than participating in weeks of development meetings [48].

Pedagogical validation evaluates the framework from both perspectives of trainer usability and trainee learning effectiveness. Construction safety trainers assess the authoring system usability using the System Usability Scale (SUS) [54] and Technology Acceptance Model constructs [54, 55]. Trainers with ⩾2 years safety training development experience complete scenario generation tasks while providing think-aloud protocols [56]. Cognitive load during authoring is assessed to evaluate the mental effort required for natural language scenario specification and generated content review, ensuring the system does not impose excessive cognitive burden on non-technical users [39]. Cognitive load measurement employs subjective rating scales capturing perceived mental demand during the authoring workflow [46].

Beyond the DKT-based adaptation described in Section 3.3, trainee evaluation incorporates Situation Awareness Global Assessment Technique (SAGAT) [57] to evaluate trainees' ability to perceive hazards (level 1), comprehend safety implications (level 2), and project potential accidents (level 3) during XR training scenarios. The simulation is periodically paused to query trainees on environmental state (e.g., "Where is the nearest fire extinguisher?"). Conduct cognitive load assessment using physiological indicators such as pupil diameter, heart rate variability to ensure adaptive scenarios maintain trainees within the "zone of proximal development" without inducing cognitive overload [58]. Knowledge transfer assessment can be meausred for application to novel hazard configuration. Delayed retention testing could happen at 2-week and 3-month intervals [59]. Sample size follows Cohen's guideliens for detecting medium effects with 80% power [60, 61].

4 Validating the Framework

The framework is intended to serve as a overview guide that practitioners and developers can follow to build their own automated content generation systems. This approach allows the framework to be validated through its practical application and its ability to streamline the transition from high-level safety requirements to interactive training modules in future implementations.

5 Conclusion

The primary limitation of this study is the current lack of full-scale implementation and empirical data to quantify the system's effectiveness. While the architecture is theoretically grounded by literature, its real-world performance will be influenced by computational costs of real-time multimodal data processing, inherent LLMs reliability in safety-critical content generation. Technical formalizations such as ontology mapping schemas for asset-logic binding remain at the architectural level, and future work must address international regulatory requirements as well as significant hardware requirements and data privacy considerations.

The primary contribution of this research is a modular conceptual framework that serves as a bridge between high-level safety regulations and automated spatial realization. Academically, the framework provides a theoretical foundation for future studies to investigate how automated authoring logic can be integrated within immersive environments to support specialized training. It establishes a research pathway for exploring how context-aware safety simulations can be generated dynamically. ractically, the framework provides a technical blueprint for XR developers to overcome the content creation bottleneck through four key contributions. (1) an LLM-to-machine executable pipeline that translates unstructured natural language instructions into structured, machine-interpretable instructions for scene orchestration; (2) a PCG engine that automatically arranges 3D assets and environmental hazards based on site-specific logic and safety constraints; (3) a DKT-driven adaptive engine that personalizes the training module by dynamically adjusting scenario difficulty to match the learner's progress; and (4) a multimodal evaluation method that combines subjective and objective training data to validate both the usability and pedagogical effectiveness of the generated content. By distinguishing between these core automated features and advanced adaptive modules, the framework offers a scalable pathway for construction firms to produce high-fidelity, regulation-compliant safety training with significantly reduced technical expertise and resource requirement.

6 Acknowledgement

The authors would like to express their sincere acknowledgement to the Discovery Grants Program (RGPIN-2023-04239) from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] A. Alsharef, A. Albert, I. Awolusi, and E. Jaselskis. Severe injuries among construction workers: Insights from OSHA's new severe injury reporting program. *Safety science*, 163: 106126, 2023.

[2] B. Y. Jeong. Occupational deaths and injuries in the construction industry. *Applied ergonomics*, 29 (5): 355-360, 1998.

[3] A. Albert, B. Pandit, and Y. Patil. Focus on the fatal-four: Implications for construction hazard recognition. *Safety science*, 128: 104774, 2020.

[4] X. Li, W. Yi, H.-L. Chi, X. Wang, and A. P. Chan. A critical review of virtual and augmented reality (VR/AR) applications in construction safety. *Automation in construction*, 86: 150-162, 2018.

[5] J. M. Davila Delgado, L. Oyedele, T. Beach, and P. Demian. Augmented and virtual reality in construction: Drivers and limitations for industry adoption. *Journal of construction engineering and management*, 146 (7): 04020079, 2020.

[6] M. D. Joe, "Instructional Design Techniques used to develop virtual reality-based Safety Training in an Industrial Environment," Walden University, 2020.

[7] S. Li, Q.-C. Wang, H.-H. Wei, and J.-H. Chen. Extended reality (XR) training in the construction industry: A content review. *Buildings*, 14 (2): 414, 2024.

[8] S. Sharma, C. Keighrey, S. Gilligan, J. Lardner, and N. Murray, "Transforming Design Reviews with XR: A No-Code Media Experience Creation Strategy for Manufacturing Design," in *Proceedings of the 2025 ACM International*

Conference on Interactive Media Experiences, 2025, pp. 30-48.
[9] S. Sharma, C. Keighrey, J. Lardner, S. Gilligan, and N. Murray, "No code XR-CAx experience creation workflows for conducting Equipment Design Reviews," in *Proceedings of the 16th ACM Multimedia Systems Conference*, 2025, pp. 322-327.
[10] M. Zahabi and A. M. Abdul Razak. Adaptive virtual reality-based training: a systematic literature review and framework. *Virtual Reality*, 24 (4): 725-752, 2020.
[11] Uptale. On-Line: https://www.uptale.io/en/ Accessed: Dec 11, 2025.
[12] Thinklink. On-Line: https://www.thinglink.com/ Accessed: Dec 11, 2025.
[13] Zapworks. On-Line: https://zap.works/ Accessed: Dec 11, 2025.
[14] J. Bräker, J. Hertel, and M. Semmann. Empowering users to create augmented reality-based solutions–deriving design principles for No-Code AR authoring tools. 2023.
[15] M.-J. Wang, C.-H. Tseng, and C.-Y. Shen, "An easy to use augmented reality authoring tool for use in examination purpose," in *IFIP Human-Computer Interaction Symposium*, 2010: Springer, pp. 285-288.
[16] C. Craven, A. Torres, B. Kapralos, D. Chandross, and A. Dubrowski, "The Development of a No-Code VR Authoring Platform for Post-Secondary Educators," in *International Conference on Extended Reality*, 2025: Springer, pp. 182-190.
[17] Motive. On-Line: https://www.motive.io/ Accessed: Dec 11, 2025.
[18] Skybox AI. On-Line: Accessed: Dec 11,2025.
[19] Inworld. On-Line: inworld.ai Accessed: Dec 11, 2025.
[20] How Ubisoft's New Generative AI Prototype Changes the Narrative for NPCs. On-Line: https://news.ubisoft.com/en-us/article/5qXdxhshJBXoanFZApdG3L/how-ubisofts-new-generative-ai-prototype-changes-the-narrative-for-npcs Accessed: Dec 11, 2025.
[21] Z. Hu *et al.*, "Scenecraft: An llm agent for synthesizing 3d scenes as blender code," in *Forty-first International Conference on Machine Learning*, 2024.
[22] W. Feng *et al.* Layoutgpt: Compositional visual planning and generation with large language models. *Advances in Neural Information Processing Systems*, 36: 18225-18250, 2023.
[23] X. Wang, H. Zhang, T. Ge, W. Yu, D. Yu, and D. Yu. Opencharacter: Training customizable role-playing llms with large-scale synthetic personas. *arXiv preprint arXiv:2501.15427*: 2025.
[24] N. Chen, X. Lin, H. Jiang, and Y. An, "Automated Building Information Modeling Compliance Check through a Large Language Model Combined with Deep Learning and Ontology," *Buildings*, vol. 14, no. 7*,* p. 1983doi: 10.3390/buildings14071983.
[25] S. V.-T. Tran *et al.* Leveraging large language models for enhanced construction safety regulation extraction. *Journal of Information Technology in Construction*, 29: 1026-1038, 2024.
[26] F. Sammour, J. Xu, X. Wang, M. Hu, and Z. Zhang. Responsible AI in construction safety: Systematic evaluation of large language models and prompt engineering. *Journal of Construction Engineering and Management*, 152 (1): 04025217, 2026.
[27] P. E. Love, W. Fang, J. Matthews, and H. Luo. Open the 'Black Box'of Large Language Models and Make Them Explainable in Construction. *IEEE Engineering Management Review*: 2026.
[28] N. A. Barriga. A short introduction to procedural content generation algorithms for videogames. *International Journal on Artificial Intelligence Tools*, 28 (02): 1930001, 2019.
[29] D. M. De Carli, F. Bevilacqua, C. T. Pozzer, and M. C. d'Ornellas, "A survey of procedural content generation techniques suitable to game development," in *2011 Brazilian symposium on games and digital entertainment*, 2011: IEEE, pp. 26-35.
[30] H. Kim, S. Lee, H. Lee, T. Hahn, and S. Kang, "Automatic generation of game content using a graph-based wave function collapse algorithm," in *2019 IEEE conference on games (CoG)*, 2019: IEEE, pp. 1-4.
[31] J. Agarwal and S. Shridevi. Procedural content generation using reinforcement learning for disaster evacuation training in a virtual 3d environment. *IEEE Access*, 11: 98607-98617, 2023.
[32] A. S. Joshi. Reinforcement Learning-Enhanced Procedural Generation for Dynamic Narrative-Driven AR Experiences. *arXiv preprint arXiv:2501.08552*: 2025.
[33] M. F. Maleki and R. Zhao, "Procedural content generation in games: A survey with insights on emerging llm integration," in *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2024, vol. 20, no. 1, pp. 167-178.
[34] M. Knowles. Adult learning processes: Pedagogy and andragogy. *Religious Education*, 72 (2): 202-211, 1977.
[35] H. Kang, S. Lee, S. Choe, D. Choi, and J. Ryu, "Designing XR-Based Training for Construction Safety Using Embodied Learning," in *Society for Information Technology & Teacher Education International Conference*, 2025: Association for the

Advancement of Computing in Education (AACE), pp. 1724-1728.

[36] L. Dawley and C. Dede, "Situated learning in virtual worlds and immersive simulations," in *Handbook of research on educational communications and technology*: Springer, 2013, pp. 723-734.

[37] L. Pietschmann, "Human-Computer Interaction in Extended Reality: Exploring the Impact of Visual Guidance on User Performance and Human Factors," 2024.

[38] D. Salcedo *et al.* Frequently used conceptual frameworks and design principles for extended reality in health professions education. *Medical science educator*, 32 (6): 1587-1595, 2022.

[39] P. V. González-Erena, S. Fernández-Guinea, and P. Kourtesis. Cognitive assessment and training in extended reality: Multimodal systems, clinical utility, and current challenges. *Encyclopedia*, 5 (1): 8, 2025.

[40] M. Meccawy. Creating an immersive XR learning experience: A roadmap for educators. *Electronics*, 11 (21): 3547, 2022.

[41] Y. Wu, G. Lee, Q. Mei, C. Mourgues, and V. Gonzalez, "Adapting Content Generation Knowledge to Extended Reality Based Training in Construction: A Systematic Literature Review," 2025.

[42] S. Sharma, C. Keighrey, S. Gilligan, J. Lardner, and N. Murray, "Transforming Design Reviews with XR: A No-Code Media Experience Creation Strategy for Manufacturing Design," presented at the *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences*, 2025.

[43] C. Piech *et al.* Deep knowledge tracing. *Advances in neural information processing systems*, 28: 2015.

[44] B. M. Velichkovsky, A. Rothert, M. Kopf, S. M. Dornhöfer, and M. Joos. Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5 (2): 145-156, 2002.

[45] Z. Wang *et al.* User-centric immersive virtual reality development framework for data visualization and decision-making in infrastructure remote inspections. *Advanced Engineering Informatics*, 57: 102078, 2023.

[46] K. Fraser, I. Ma, E. Teteris, H. Baxter, B. Wright, and K. McLaughlin. Emotion, cognitive load and learning outcomes during simulation training. *Medical education*, 46 (11): 1055-1062, 2012.

[47] A. F. Borthick, D. R. Jones, and S. Wakai. Designing Learning Experiences within Learners' Zones of Proximal Development (ZPDs): Enabling Collaborative Learning On - Site and Online. *Journal of Information Systems*, 17 (1): 107-134, 2003.

[48] S. Chidambaram *et al.*, "EditAR: A Digital Twin Authoring Environment for Creation of AR/VR and Video Instructions from a Single Demonstration," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 17-21 Oct. 2022 2022, pp. 326-335, doi: 10.1109/ISMAR55827.2022.00048.

[49] T. Hu, T. Du, Z. Qu, and M. Gorlatova, "XR Reality Check: What Commercial Devices Deliver for Spatial Tracking," in *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2025: IEEE, pp. 932-942.

[50] K. Jasani. Performance optimization in VR applications: QA's role. *IJAIDR-Journal of Advances in Developmental Research*, 15 (1): 2024.

[51] F. Vona, M. Warsinke, T. Kojić, J.-N. Voigt-Antons, and S. Möller, "User-centric evaluation methods for digital twin applications in extended reality," in *2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*, 2025: IEEE, pp. 142-146.

[52] L. C. Bertens *et al.* Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. *PLoS medicine*, 10 (10): e1001531, 2013.

[53] S. Rajendran and J. A. Gambatese. Development and initial validation of sustainable construction safety and health rating system. *Journal of construction engineering and management*, 135 (10): 1067-1075, 2009.

[54] J. Brooke. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189 (194): 4-7, 1996.

[55] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 13 (3): 319-340, 1989.

[56] C. Lewis. Using the'thinking-aloud'method in cognitive interface design. *Research Report RC9265, IBM TJ Watson Research Center*: 1982.

[57] C.-Y. Cheng and B. Esmaeili, "Situation awareness study in the construction industry: A systematic review," in *Construction Research Congress 2024*, 2024, pp. 843-853.

[58] Z. Wang, Z. Jiang, and A. Blackman. Linking emotional intelligence to safety performance: The roles of situational awareness and safety training. *Journal of safety research*, 78: 210-220, 2021.

[59] T. Sitzmann and J. M. Weinhardt. Training engagement theory: A multilevel perspective on the effectiveness of work-related training. *Journal of Management*, 44 (2): 732-756, 2018.

[60] J. Cohen, *Statistical power analysis for the behavioral sciences*. routledge, 2013.