# Stepping Into Stories: Envisioning a Generative AI Pipeline to Create Story-based VR Reading Environments

Vitali, A.; Schneegass, C.; Dingler, Tilman

**DOI**
10.18420/muc2025-mci-ws09-144

**Publication date**
2025

**Document Version**
Final published version

**Citation (APA)**
Vitali, A., Schneegass, C., & Dingler, T. (2025). *Stepping Into Stories: Envisioning a Generative AI Pipeline to Create Story-based VR Reading Environments*. Mensch und Computer 2025 , Chemnitz, Germany. https://doi.org/10.18420/muc2025-mci-ws09-144

# Stepping Into Stories: Envisioning a Generative AI Pipeline to Create Story-based VR Reading Environments

Alice Vitali
Delft University of Technology
Delft, Netherlands
a.vitali@tudelft.nl

Christina Schneegass
Delft University of Technology
Delft, Netherlands
c.schneegass@tudelft.nl

Tilman Dingler
Delft University of Technology
Delft, Netherlands
t.dingler@tudelft.nl

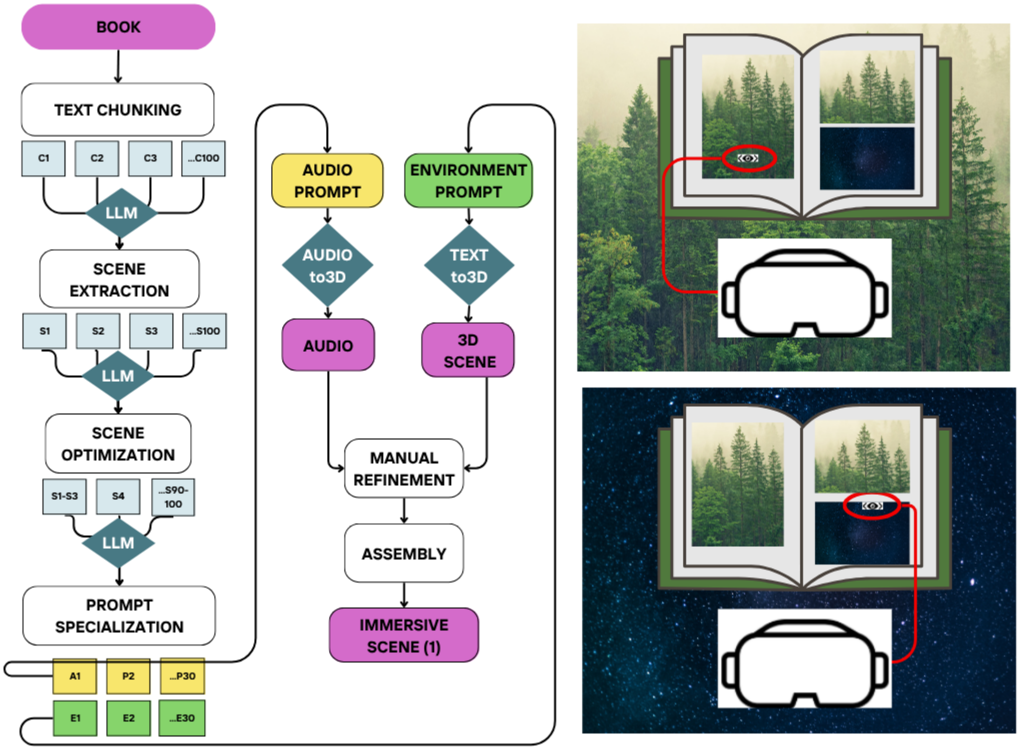

**Figure 1: Flowchart of the generation of story-based immersive reading environments from text. The GenAI pipeline (left) transforms fictional text into dynamic VR environments that adapt to story progression (right).**

## Abstract

*Where* you read matters—so what if you could read *literally inside* your book? Reading in Virtual Reality (VR) has been shown to support deep immersion and narrative engagement. We argue that Generative Artificial Intelligence (GenAI) can significantly expand what is currently possible in VR reading by dynamically producing story-driven environments from text. In this vision paper, we present a pipeline that combines recent advances in language, audio, and 3D scene generation to allow automatically augmenting fiction reading with environmental backdrops. As the reader progresses, these AI-generated environments transition in real time, acting as cognitive props for visual imagery. We also describe a user evaluation, discuss limitations, and address implications and open questions raised by the proposed approach.

## CCS Concepts

• **Computing methodologies** → *Artificial intelligence.*

## Keywords

Generative Artificial Intelligence, Virtual Reality, Reading, Large Language Models, Text-to-3D

*Mensch und Computer 2025 – Workshopband, Gesellschaft für Informatik e.V., 31. August - 03. September 2025, Chemnitz, Germany*

https://doi.org/10.18420/muc2025-mci-ws09-144

## 1 Introduction

Imagine reading your copy of The Hitchhiker's Guide to the Galaxy [1] while sitting in the Heart of Gold's white cabin, covered in control panels and screens, hearing the hum of malfunctioning machinery. This is how recent advances in the fields of Generative

Artificial Intelligence (GenAI) and Virtual Reality (VR) could be leveraged to create an embodied narrative experience that enhances the pleasure of reading.

In today's era of technological hyper-stimulation and ubiquitous screens, fewer young people read in their free time [5], and reading strategies are shifting towards skimming and skipping, rather than deep engagement with text [12]. Yet, the benefits that come from long-form, immersed reading are numerous and well-documented—from improved literary skills to enhanced cognitive function and empathy [4].

To address the shift away from pleasure reading, researchers are exploring the potential of immersive technologies. VR reading environments [10, 11, 15, 16] have shown promise: they offer a distraction-free setting that supports sustained attention, making reading more appealing (especially to younger audiences). Moreover, environmental features tied to the story content can significantly enhance reader immersion and enjoyment, fostering a deeper connection to the narrative by creating a perceived shift in embodiment from the real world to the virtual world [3, 11].

However, how to design effective environments that dynamically reflect a story's content is still an open question. Although earlier studies relied on pre-existing book illustrations [3], we argue that the potential of generative models remains under-leveraged in this context. While AI already powers key VR capabilities (e.g., computer vision for live tracking and gesture recognition), text-to-audio and text-to-3D models can now produce high-quality content with minimal human intervention [17, 18]. By reducing the burden of modeling from scratch, generative models enable quicker prototyping for digital storytellers without technical expertise, facilitating early user testing and iterative refinements. Moreover, genAI enables the creation of highly personalized experiences where content adapts to readers' individual engagement patterns in real-time—for instance, by providing more support to distracted readers, or personalized memory cues after an interruption. Finally, this generative approach scales effectively to extensive long-form content that would otherwise be prohibitively expensive to adapt manually.

In this paper, we propose a GenAI pipeline to generate story-aligned immersive reading environments, with the goal of literally transporting the reader into the story's setting. Our aim is to demonstrate how this vision could be technically realized by combining recent advances in multi-modal generative AI. We also seek to spark discussion around the opportunities and concerns raised by applying such technology at the intersection of reading and VR.

## 2 Story-to-VR Pipeline

Our proposed pipeline transforms text into a series of immersive environments through five sequential steps, as outlined below; see Figure 1 (left) for the flowchart of the pipeline. We incorporate instruction-driven 3D modeling frameworks, like 3D-GPT [20], that use LLMs to orchestrate procedural generation tasks, reducing the technical complexity of 3D content creation. While generative AI will be used to assist and fasten content generation, human refinement and iterative testing will be fundamental throughout the process (e.g., to select the optimal number and type of transitions, scenery triggers, and overall design).

(1) **Text chunking** The input story is segmented using document-specific or semantic chunking to fit into the LLM's context window while preserving narrative coherence [2].

(2) **Scene Extraction** A Large Language Model (LLM) processes each chunk to both extract and infer environmental cues (e.g., location, background objects, weather, time of the day, etc.), generating a detailed scene description. For example, from a chunk of text describing the Heart of Gold's white cabin, the LLM could generate a description mentioning the inside of a spacecraft with metallic surfaces, monitors, and panels.

(3) **Scene Optimization** Initial scenes are refined through two operations: consecutive scenes with similar environmental settings are merged to avoid redundant transitions (e.g., multiple scenes in the same location), while scenes spanning multiple distinct locations are split to ensure each environment change aligns with narrative progression (e.g., a character moving from forest to castle). Different natural language processing (NLP) techniques may be suitable for this task, from a simple keyword-based approach to semantic similarity computed with word embeddings.

(4) **Prompt Specialization** Each description is transformed into specialized prompts optimized for different modalities and content types through automatic prompt generation [23]. For example, starting from the Heart of Gold's white cabin description, specialized prompts might include:
   - Visual prompt: "Generate a futuristic spacecraft interior with white metallic panels, multiple glowing control screens, and a sterile but lived-in atmosphere"
   - Audio prompt: "Create ambient spacecraft sounds: low electrical humming, occasional beeping from malfunctioning systems, distant mechanical whirring"
   - Atmospheric prompt: "Convey a sterile yet chaotic mood through lighting: harsh fluorescent overhead lights with flickering, casting sharp shadows on control surfaces"

   The prompts above were generated using Claude [1] for illustrative purposes.

(5) **Scene Generation, Refinement, and Assembly** For each scene, the specialized prompts from the previous step become input to their respective models; for instance, an instruction-driven 3D modeling framework (e.g., 3D-GPT) produces the **3D model** of the environment, and a text-to-audio model creates **soundscapes**. Maintaining consistency across scenes would require memory mechanisms to preserve recurring environmental elements, ensuring smooth transitions. Models' outputs are then assembled into functional 3D scenes through spatial layout, lighting integration/audio specialization and interaction mapping—enabling a default, fully automated story adaptation. Optionally, outputs may be manually refined using 3D sculpting tools (e.g., Blender) and audio editing software to achieve higher quality or more specific creative vision, supporting human-in-the-loop enhancement if resources permit.

The sequence of scenes generated through the pipeline is ultimately compressed and loaded into the VR reading application. While real-time generation during reading would be ideal, current

[1] Claude AI, https://claude.ai/

model inference times would require pre-generating content to ensure a smooth reading experience.

As the user reads, the system tracks reading progress to trigger environmental changes. For traditional text presentation, this can be achieved by monitoring the scroll-bar position or current page location. In the case of Rapid Serial Visual Presentation (RSVP)—where words appear sequentially at a controlled pace—the system can track progress through line or word count. Later implementations could consider using eye-tracking data to trigger the transition to the next environment at the right moment (see Figure 1 (right)). Additionally, users could modulate environmental features in real-time (e.g., adjusting lighting intensity, ambient sound levels, or visual complexity) to match their reading preferences.

## 2.1 Proposed Evaluation

To succeed in creating AI-generated environments that enhance reading without overwhelming the reader, we propose a within-subject study design comparing conditions that vary key hyperparameters such as *environment transition frequency* (static, chapter-level, paragraph-level) and *environmental abstraction levels* (concrete, moderate, abstract). This user-evaluation framework is not intended as a single assessment at the end of the transformation of text into complete scenes, but serves as an integral component throughout the pipeline development stages—providing human guidance for prompt engineering, parameter tuning, and system refinement based on reader response patterns.

*Measures.* The Story World Absorption Scale (SWAS) [9] will assess narrative absorption across four dimensions (attention, transportation, emotional engagement, mental imagery). Additional measures include reading engagement, comprehension retention, and preference ratings. Eye-tracking data will analyze readers' gaze patterns to assess whether environments support text focus or cause distraction.

## 2.2 Implementation Challenges

Effective implementation of this approach requires addressing several open challenges across different stages of the pipeline:

- **Text segmentation**: Stories may lack explicit environmental descriptions or describe imagined/remembered environments in a non-chronological order. Internal monologues, non-physical spaces or abstract passages present an additional challenge for the automatic generation of VR environments.
- **Scene consistency**: For recurring locations to be consistent throughout the narrative, the system must leverage a memory system; this would ensure that readers experience familiar locations as coherent, recognizable spaces rather than randomly regenerated environments.
- **Multi-modal Coordination**: Generated 3D environments, ambient audio, and interactive elements must semantically align and complement each other, as mismatched outputs could break immersion entirely.

Additionally, the optimal balance between technological support and potential distraction remains to be empirically determined— as too few immersive features may fail to engage readers, while too many could overwhelm them. This tension spans multiple dimensions: transition frequency (e.g., chapter-level, paragraph-level, or sentence-level), degree of interactivity (from passive observation to active manipulation), and environmental complexity (from minimalist to detailed). Finding the optimal configuration across these dimensions requires understanding how readers respond to varying levels of environmental stimulation in their reading experience.

# 3 Ethical implications and Open Questions

When developing GenAI applications, we must acknowledge several known concerns and ethical implications—from the impact of these models on the creative industry and the environment, to unresolved legal questions about ownership and authorship of the generated outputs, as well as bias amplification and misinformation [7]. While some of these risks can be partially mitigated by using open-source models [6] instead of proprietary ones, other concerns remain; for instance, training datasets frequently contain copyrighted content without the original creators' knowledge or consent [22]. Additionally, readers, and children in particular, risk being exposed to inappropriate generated content, such as graphic sexual content or horror scenes. Despite their necessity, moderation mechanisms [14] may limit the literary expression of certain genres; they may also be difficult to apply in dynamic, customised environments.

Our Story-to-VR pipeline raises an additional open question about the integration of GenAI in creative domains: how can we balance humanity and technology?—with humanity entailing creativity [21] and imagination. This tension is evident in the reading application domain on both the creation and consumption stages. In the creation process, this includes determining how the original creators (e.g., a book's author) should be involved when augmenting their work through AI, considering there is a possibility of altering their original intent. At the consumption stage, the tension centers on the reader. In traditional reading, immersion relied on mental abilities to construct fictional worlds ("phenomenological immersion"), while immersive reading environments shape experience through technological features ("technological immersion") [13], potentially reducing the active role of imagination. The risk is that readers become passive consumers rather than active participants in imaginative world-building. The challenge lies in leveraging technology to enhance this inherently creative process, rather than substituting the reader's imagination. This raises critical questions about user agency: how much control should readers maintain over their virtual environment? Should the system generate fixed interpretations of scenes, or provide customizable elements that readers can modify to match their personal vision? Where can the line be drawn between supporting reading and altering its very essence? These design decisions also open the door to potentially manipulative dark patterns—interface choices that subtly influence user behaviour for profit, data collection, or to steer attention [8]; for instance, reading environments could be designed to foreground certain narratives or fixed interpretations. Moreover, behind the pursuit of enhanced reading engagement lies the threat of technological addiction [19]; currently, we still lack a clear understanding of the long-term cognitive effects of immersive reading.

Although largely unanswered, these questions of how to balance humanity and automation must remain central when designing AI systems across all domains, to avoid a future of machine dependence while still striving for meaningful human-AI cooperation. For readers, this means supporting immersion and imagery without distracting or overwhelming them, and avoiding transforming a book into a movie—or worse, making imagination obsolete.

## 4 Conclusion

In this paper, we propose our vision of how recent advances in GenAI can be leveraged together with immersive technologies to create immersive, embodied reading experiences–literally transporting readers inside the story world. Through our five-step pipeline, we demonstrate how multi-modal generative AI models could support the creation of virtual backdrops, becoming a tool for researchers and digital artists. Ultimately, the aim is to strike a balance between technological and phenomenological immersion, leveraging technology to help readers engage in deep, sustained long-form reading. Beyond the context of reading, this vision aims to spark general discussion about the technical possibilities that lie ahead of us, but also the challenges of balancing automation with humanity, creativity, and imagination.

## References

[1] Douglas Adams. 1979. *The Hitchhiker's Guide to the Galaxy.* Pan Books, London.

[2] Analytics Vidhya. 2024. *15 Chunking Techniques to Build Exceptional RAG Systems.* Analytics Vidhya. https://www.analyticsvidhya.com/blog/2024/10/chunking-techniques-to-build-exceptional-rag-systems/ Accessed: 2025-06-08.

[3] Eric Bahna and Robert J. K. Jacob. 2005. Augmented reading: presenting additional information without penalty. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1909–1912. https://doi.org/10.1145/1056808.1057054

[4] Christina Clark and Kate Rumbold. 2006. *Reading for Pleasure: A Research Overview.* Technical Report. National Literacy Trust. https://files.eric.ed.gov/fulltext/ED496343.pdf Accessed July 2025.

[5] Aimee Cole, Brown Ariadne, Christina Clark, and Irene Picton. 2022. *Children and young people's reading engagement in 2022: Continuing insight into the impact of the Covid-19 pandemic on reading.* Technical Report. National Literacy Trust.

[6] Francisco Eiras, Aleksandar Petrov, Bertie Vidgen, Christian Schroeder, Fabio Pizzati, Katherine Elkins, Supratik Mukhopadhyay, Adel Bibi, Aaron Purewal, Csaba Botos, Fabro Steibel, Fazel Keshtkar, Fazl Barez, Genevieve Smith, Gianluca Guadagni, Jon Chun, Jordi Cabot, Joseph Imperial, Juan Arturo Nolazco, Lori Landay, Matthew Jackson, Phillip H. S. Torr, Trevor Darrell, Yong Lee, and Jakob Foerster. 2024. Risks and Opportunities of Open-Source Generative AI. arXiv:2405.08597 [cs.LG] https://arxiv.org/abs/2405.08597

[7] Ziv Epstein, Aaron Hertzmann, the Investigators of Human Creativity, Memo Akten, Hany Farid, Jessica Fjeld, Morgan R. Frank, Matthew Groh, Laura Herman, Neil Leach, Robert Mahari, Alex Sandy Pentland, Olga Russakovsky, Hope Schroeder, and Amy Smith. 2023. Art and the science of generative AI. *Science* 380, 6650 (2023), 1110–1111. https://doi.org/10.1126/science.adh4451

[8] Thomas Kollmer and Andreas Eckhardt. 2023. Dark Patterns. *Business & Information Systems Engineering* 65, 3 (2023), 201–208. https://doi.org/10.1007/s12599-022-00783-7

[9] Moniek M Kuijpers. 2023. Story World Absorption Scale. https://doi.org/10.17605/OSF.IO/ZF439

[10] Nikola Kunzova and Daniel Echeverri. 2023. Bookwander: From Printed Fiction to Virtual Reality—Four Design Approaches for Enhanced VR Reading Experiences. In *Interactive Storytelling*, Lissa Holloway-Attaway and John T. Murray (Eds.). Springer Nature Switzerland, Cham, 309–328.

[11] Anežka Kuzmičová. 2016. Does it Matter Where You Read? Situating Narrative in Physical Environment. *Communication Theory* 26, 3 (2016), 290–308. https://doi.org/10.1111/comt.12084

[12] Ziming Liu. 2005. Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation* 61, 6 (2005), 700––712. https://doi.org/10.1108/00220410510632040

[13] A. Mangen. 2008. Hypertext fiction reading: Haptics and immersion. *Journal of Research in Reading* 31 (11 2008), 404–419. https://doi.org/10.1111/j.1467-9817.2008.00380.x

[14] P. Pardhi. 2025. Content Moderation of Generative AI Prompts. *SN Computer Science* 6 (2025), 329. https://doi.org/10.1007/s42979-025-03864-y

[15] Federico Pianzola and Luca Deriu. 2021. StoryVR: A Virtual Reality App for Enhancing Reading. In *Methodologies and Intelligent Systems for Technology Enhanced Learning, 10th International Conference. Workshops*, Zuzana Kubincová, Loreto Lancia, Elvira Popescu, Minoru Nakayama, Vittorio Scarano, and Ana B. Gil (Eds.). Springer International Publishing, Cham, 281–288.

[16] Federico Pianzola and Luca Deriu. 2021. StoryVR: A Virtual Reality App for Enhancing Reading. In *Methodologies and Intelligent Systems for Technology Enhanced Learning, 10th International Conference. Workshops*, Zuzana Kubincová, Loreto Lancia, Elvira Popescu, Minoru Nakayama, Vittorio Scarano, and Ana B. Gil (Eds.). Springer International Publishing, Cham, 281–288.

[17] Arjun T. Prakash, Arjun K. S., Shijo Shaji, Sainath Raghunath, and Rekha K. S. 2025. A Survey on Text to 3D Model Generator. *International Journal for Research Trends and Innovation* 10, 1 (jan 2025), a585. https://www.ijrti.org/viewpaper.aspx?paperid=IJRTI2501071 IJRTI2501071.

[18] Jaspreet Singh, Gurpreet Singh, Rajat Verma, and Chander Prabha. 2023. Exploring the Evolving Landscape of Extended Reality (XR) Technology. In *2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*. IEEE, Bangalore, India, 1–6. https://doi.org/10.1109/SMARTGENCON60755.2023.10442251

[19] Gary W Small, Jay Lee, Ariel Kaufman, Javad Jalil, Prabha Siddarth, Harsha Gaddipati, Thomas D Moody, and Susan Y Bookheimer. 2020. Brain health consequences of digital technology use. *Dialogues in Clinical Neuroscience* 22, 2 (2020), 179–187. https://doi.org/10.31887/DCNS.2020.22.2/gsmall

[20] Chunyi Sun, Junlin Han, Weijian Deng, Xinlong Wang, Zishan Qin, and Stephen Gould. 2024. 3D-GPT: Procedural 3D Modeling with Large Language Models. arXiv:2310.12945 [cs.CV] https://arxiv.org/abs/2310.12945

[21] Yiren Xu. 2025. Balancing Creativity and Automation: The Influence of AI on Modern Film Production and Dissemination. arXiv:2504.19275 [cs.CY] https://arxiv.org/abs/2504.19275

[22] K.-Q. Zhou and H. Nabus. 2023. The Ethical Implications of DALL-E: Opportunities and Challenges. *Mesopotamian Journal of Computer Science* 2023 (2023), 17–23.

[23] Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. 2023. Large Language Models Are Human-Level Prompt Engineers. arXiv:2211.01910 [cs.LG] https://arxiv.org/abs/2211.01910