PODSAGE: Story-Driven AI for Enhanced Learning, Comprehension, and Retention

Ismail Ben seddik
isbens@iu.edu

Abstract

PODSAGE is an AI-driven educational storytelling platform that transforms complex concepts into engaging, interactive narratives. Using a fine-tuned LLaMA 3.3 model, PODSAGE produces multimodal stories with credible citations, adaptive personalization, and support for neurodivergent learners. Unique features include interactive narrative choices, persona-aware delivery, and a community content studio. Evaluation combines human and LLM-based scoring. PODSAGE demonstrates a scalable, ethical approach to enhancing comprehension and retention through the science of storytelling.

1 Introduction

Storytelling has long been recognized as one of the most powerful tools for communicating science, teaching professional knowledge, and fostering deep learning among non-experts and beginners (Green and Brock, 2000; Graesser et al., 2004; Shift, 2023). Neuroscience and educational research have shown that well-crafted stories activate multiple brain regions, enhance attention, and significantly improve retention and comprehension (Shift, 2023; Speer et al., 2009; Trabasso and van den Broek, 1985). Yet, despite these benefits, essential concepts in fields such as law, information technology, and finance remain inaccessible to many due to their inherent complexity and a lack of engaging educational materials (Jiang et al., 2024; Graesser et al., 2004).

This challenge is further exacerbated by a marked decline in attention spans and retention rates among children and young adults over the past decade. Recent studies indicate that over 30% of students struggle to retain information from traditional classroom methods, while the rise of short-form, low-effort digital content (e.g., TikTok, YouTube Shorts) is rewiring brains for quick dopamine bursts rather than deep, sustained thinking (Shift, 2023). The result is a growing gap between the accessibility of fun, brainless content and the cognitive demands of mastering complex, real-world knowledge.

Storytelling, when designed with emotional intelligence and pedagogical intent, can bridge this gap. Emotionally resonant narratives not only deliver dopamine and delight but also spark curiosity, foster empathy, and create durable memories (Green and Brock, 2000; Speer et al., 2009; Kensinger and Schacter, 2006). However, the scalable implementation of storytelling-based education is severely limited by the high costs and expertise required to craft such narratives, especially in specialized domains (Jiang et al., 2024).

Recent advances in large language models (LLMs) have opened new avenues for automating the creation of high-quality, contextually rich stories (Jiang et al., 2024; Graesser et al., 2004; August et al., 2024). Notably, MIT's study on leveraging LLMs for legal education found that LLM-generated stories not only improved comprehension and retention but also made abstract legal concepts more relatable and engaging for non-experts (Jiang et al., 2024).

In this work, we introduce PODSAGE, an application that leverages LLMs to transform complex ideas into engaging, well-narrated stories optimized for multi-modal learning. PODSAGE offers interactive, personalized content generation in real time, featuring custom playlists, smart topic discovery, and story curation based on cognitive and learning science principles. Our system uses Llama 3.3 (8B parameters) for fast inference, fine-tuned on a curated dataset of high-quality stories generated by GPT-4.1, and incorporates Retrieval-Augmented Generation (RAG) for accurate concept definitions and peer-reviewed references. To ensure quality and trustworthiness, PODSAGE includes expert and moderator review

workflows as well as an LLM-based judge for evaluating story accuracy, structure, and educational effectiveness.

The main contributions are:

- A scalable and user-centric framework for generating emotionally engaging, pedagogically effective stories from complex concepts using LLMs.
- A demonstration of how fine-tuning a lightweight LLM on a curated dataset from a more powerful model (GPT-4.1) enables high-quality, efficient story generation.
- Providing a platform for human and LLM-based evaluation, supporting continuous improvement and trust in educational storytelling.

2 Related Work

Storytelling has long been recognized as a powerful pedagogical tool for conveying complex concepts in education, science, and professional domains (Busselle and Bilandzic, 2009, 2008; Rapaport et al., 1989; Papadimitriou, 2003; Abbott, 2020). Narratives, or stories, are sequential depictions of actions and events that facilitate understanding and retention, particularly for non-experts and beginners. Research in cognitive science and education has demonstrated that storytelling enhances engagement, comprehension, and memory by leveraging narrative transportation, emotional resonance, and thematic coherence (Green and Brock, 2000; Speer et al., 2009; Trabasso and van den Broek, 1985; Graesser et al., 2004; Kensinger and Schacter, 2006).

In legal education, storytelling is especially prominent through the use of "fact-patterns" to illustrate abstract doctrines (Papadimitriou, 2003). However, the scalable creation of such narratives has traditionally required significant expert involvement and resources. The recent MIT study by Jiang et al. (Jiang et al., 2024) presents a novel application of large language models (LLMs) for generating legal stories. Their LegalStories dataset, consisting of 295 legal doctrines paired with LLM-generated stories and questions, demonstrates that LLM-generated narratives significantly improve comprehension and retention among non-experts compared to definitions alone. This work highlights the promise of LLMs in democratizing access to specialized knowledge through storytelling.

Automatic story generation has been a longstanding challenge in natural language processing (NLP). Early approaches, including those based on transformer architectures, often struggled with coherence, character consistency, and narrative structure (Alabdulkarim et al., 2021; Khalil et al., 2020). The advent of LLMs such as ChatGPT has led to significant improvements in story quality, coherence, and engagement (Xie et al., 2023; Zimmerman et al., 2022; Xu et al., 2020). Recent frameworks have harnessed LLMs to generate educational stories that demystify complex legal, health, and scientific concepts for non-experts, creating dynamic knowledge bases and exemplars for continuous learning (Engineering, 2024; Landrum et al., 2023).

Despite these advances, relatively few studies have focused on the use of LLM-generated stories specifically for educational purposes. Valentini et al. (Valentini et al., 2023) explored LLMs to create age-appropriate stories for children, while other works have examined the cognitive and affective mechanisms underlying storytelling in science and STEM education (Landrum et al., 2023). Research also shows that second-person narratives and advanced narrative techniques can enhance emotional and aesthetic involvement, further supporting learning and engagement (Mildorf, 2016; Rembowska-Płuciennik, 2022).

Beyond text, AI-driven podcast generators such as Wondercraft and Podcast-LLM have begun to transform written content into engaging audio narratives, leveraging text-to-speech and conversational AI to expand accessibility and reach (Wondercraft, 2024; Dempsey, 2024; ElevenLabs, 2024). However, these systems typically focus on content repurposing rather than the generation of pedagogically optimized stories.

Recent work has also explored fine-tuning smaller LLMs on high-quality outputs from larger models to enable efficient, domain-adapted story generation (Ivison et al., 2024). Cognitive and learning science research continues to inform best practices for narrative structure, pacing, and emotional engagement, further guiding the development of effective educational storytelling systems (Green and Brock, 2000; Graesser et al., 2004; Speer et al., 2009; Trabasso and van den Broek, 1985; Kensinger and Schacter, 2006; Yao et al., 2011; Abraham et al., 2012).

To my knowledge, PODSAGE is the first system

to combine fine-tuned LLMs and multimodal delivery to produce interactive, high-quality stories that make complex concepts accessible, engaging, and memorable for diverse audiences.

3 Dataset & Methods

3.1 Approach & Justification

3.1.1 Architecture

PODSAGE is designed as a scalable, cross-platform educational storytelling platform, leveraging a robust microservices architecture to ensure modularity, flexibility, and seamless user experience across devices (Figure 1). The frontend is implemented in Flutter, enabling rapid deployment to web, Android, and iOS platforms from a single codebase.

User management and authentication are handled by a dedicated Spring Boot service, providing secure and efficient role-based access and profile management. Machine learning services-including the story generation engine, personalization pipeline, and recommender system-are built using Django and FastAPI, chosen for their maturity in production-grade ML deployment and ease of integration with Python-based LLM frameworks. Real-time sentiment analysis and user behavior tracking inform these services, supporting dynamic content adaptation and personalized recommendations, consistent with best practices in adaptive learning and affective computing.

Audio streaming, including on-demand podcast delivery, is managed by a Node.js service, ensuring low-latency, scalable media serving. Data management leverages a polyglot persistence strategy: PostgreSQL for structured user and content metadata, MongoDB for flexible story and interaction logs, Redis for high-speed caching, and cloud object storage buckets for media assets. This approach aligns with modern best practices for scalable, data-intensive applications.

PODSAGE Studio, the content creation and editing suite, integrates with ElevenLabs for natural-sounding text-to-speech and story narration, and incorporates AI-powered sound and music generation that is automatically embedded into stories. This multimodal approach is grounded in cognitive and learning science research showing that vivid, emotionally engaging audio-visual content enhances comprehension and retention.

The entire platform is deployed on Google Cloud Platform, leveraging Google Kubernetes Engine for container orchestration, cloud load balancers for traffic management, and managed services for database and storage. This ensures high availability, fault tolerance, and seamless scaling to accommodate fluctuating user demand.

This architectural design is justified by the need for real-time, personalized, and multimodal educational storytelling at scale. The microservices approach enables independent scaling and rapid iteration of core components, while the use of state-of-the-art ML frameworks and cloud-native infrastructure ensures both efficiency and extensibility. Integrating advanced TTS and audio generation aligns with evidence from neuroscience and educational psychology that multisensory, emotionally engaging narratives are more effective for learning complex concepts. The choice of cloud deployment and polyglot persistence further supports the platform's goals of accessibility, reliability, and adaptability for diverse educational contexts.

3.1.2 Story Generation

PODSAGE's story generation module is designed to produce engaging, accurate, and pedagogically effective educational stories across a wide range of domains. To balance high-quality output with fast, scalable inference, the open-source LLaMA 3.3 model (8B parameters) is used, which is lightweight enough for real-time deployment yet powerful enough for nuanced narrative generation. The model was selected for its effectiveness in text generation and its open licensing, facilitating both research and production use.

To ensure the highest quality stories, a two-stage pipeline is employed. First, a dataset of 22 educational stories is curated, covering the 11 core themes defined by PODSAGE (see Appendix), with two stories per theme. Each story is generated by a stronger LLM (ChatGPT 4.1), leveraging its superior narrative and pedagogical abilities. This approach follows recent work showing that fine-tuning smaller models on outputs from more capable LLMs can substantially improve downstream generation quality (Jiang et al., 2024; Xie et al., 2023). Parameter-efficient fine-tuning is then performed on LLaMA 3.3 using this curated dataset, focusing on narrative structure, conceptual clarity, and educational value.

For inference, a rigorously engineered zero-shot prompt (see Appendix) is used, crafted and iteratively refined based on cognitive science, neuroscience, and educational research, ensuring that

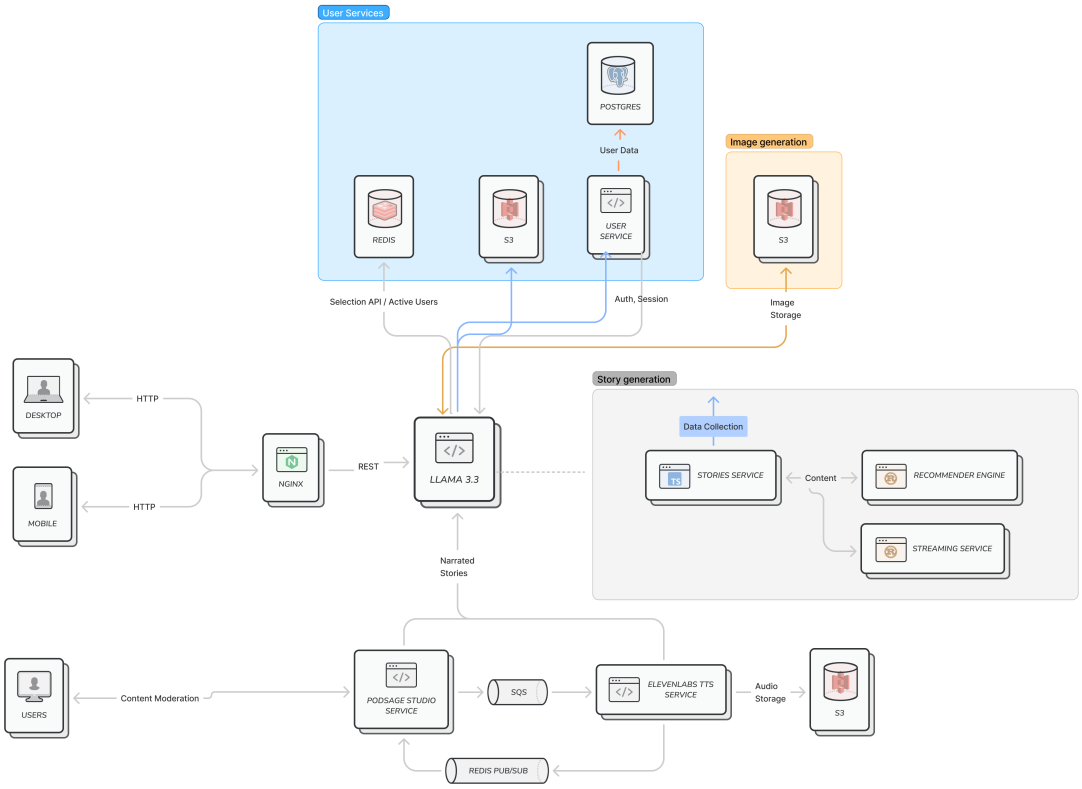

Figure 1: PODSAGE App: Tentative Technical Architecture.

each generated narrative maximizes engagement, comprehension, and retention.

Research consistently shows that stories with a clear narrative arc-beginning, middle, and end-facilitate comprehension and memory by providing a logical framework for organizing information (Waterford.org, 2024; Agency, 2024; Abbott, 2020). Story structure elements such as character, setting, conflict, and resolution help learners of all ages understand and retain complex content (Waterford.org, 2024). The use of storyboards and story maps further reinforces structural comprehension.

Relatable Characters and Dialogue Well-developed, relatable characters and authentic dialogue activate brain regions associated with social cognition and empathy, such as the medial prefrontal cortex, making abstract concepts more tangible and memorable (Speer et al., 2009). Including characters that readers can empathize with and realistic conversations humanizes complex ideas and enhances emotional engagement.

Causal Links and Thematic Unity Maintaining causal coherence-clear cause-and-effect relationships between events-supports the brain's ability to encode and retrieve information effectively (Trabasso and van den Broek, 1985; van den Broek et al., 2001). Thematic unity and logical progression are essential for comprehension, especially when conveying challenging or abstract topics.

Vivid Descriptions and Sensory Language Vivid, sensory-rich descriptions activate perceptual brain areas, including the visual cortex, and foster mental imagery, which has been shown to improve understanding and retention (Yao et al., 2011). The prompt encourages the use of such language to make abstract ideas concrete.

Emotional Engagement and Surprise Emotionally charged stories, especially those with unexpected endings or plot twists, activate the amygdala and hippocampus, enhancing memory consolidation and attention (Kensinger and Schacter, 2006; Kang et al., 2009). Incorporating emotional hooks and surprise elements in the narrative makes the content more memorable and impactful.

Analogies, Metaphors, and Real-World Examples Analogies and metaphors bridge the gap between unfamiliar concepts and familiar experiences, activating the inferior frontal gyrus and an-

gular gyrus for intuitive understanding (Thibodeau and Boroditsky, 2011; Agency, 2024). Real-world examples and visual metaphors help audiences grasp complex scientific or technical ideas more easily.

Audience Tailoring and Plain Language Effective storytelling requires adapting to the audience's background, interests, and level of understanding (Agency, 2024; Publishing, 2017). Using plain language, short sentences, and familiar vocabulary ensures that stories are accessible without oversimplifying the content.

Active and Dialogic Storytelling Interactive and dialogic storytelling-inviting the reader to participate or reflect-aligns with constructivist learning theories and has been shown to enhance problem-solving and conceptual understanding (Graesser et al., 2004). Prompts may include questions or moments for reflection to maintain engagement.

Pacing, Pauses, and Cognitive Load Thoughtful pacing and strategic pauses allow for cognitive processing and information consolidation, which is critical when narrating complex or concept-heavy stories (Carpenter and Just, 1975). The prompt encourages moments for reflection to aid learning.

Visual and Multimodal Aids Incorporating visual elements (e.g., illustrations, storyboards) and multimodal content (audio, music, SFX) supports dual coding theory, catering to both visual and auditory learners and further enhancing retention (Yao et al., 2011; Publishing, 2017).

Justification By integrating these evidence-based elements into the story generation prompt, PODSAGE ensures that each story is not only engaging and accessible but also scientifically optimized for learning and memory. This approach is supported by recent neuroscientific findings showing that storytelling activates and synchronizes multiple brain regions, leading to deeper understanding and longer-lasting retention (Saito et al., 2019; Agency, 2024; Green and Brock, 2000; Abraham et al., 2012).

To ensure factual accuracy and up-to-date content, Retrieval-Augmented Generation (RAG) is integrated into the story generation pipeline. RAG retrieves topic definitions and relevant source documents from a vector database and injects them into the LLM prompt, grounding the generated stories in authoritative, current information (Zilliz, 2024; Shelf.io, 2024; Cloud, 2025). This approach not only reduces hallucinations but also enables the model to cite credible references at the end of each story, improving transparency and trustworthiness (Zilliz, 2024; Shelf.io, 2024).

3.1.3 Content Generation Pipeline

At the core of the content generation pipeline (Figure 2) is LLaMA 3.3 (8B parameters), which generates educational stories, each paired with a title and summary, forming the narrative foundation for subsequent multimodal content creation.

To enhance engagement and support dual coding theory in learning, PODSAGE uses DeepFloyd IF, a leading text-to-image model, to generate visually compelling story covers and illustrations based on story summaries. This integration of visual elements has been shown to improve memory retention and learner motivation (Yao et al., 2011; Abraham et al., 2012).

For audio narration, PODSAGE integrates ElevenLabs' advanced text-to-speech (TTS) and generative audio tools (ElevenLabs, 2024; Voicebot.ai, 2024; ElevenLabs, 2023). ElevenLabs narrates the story with natural, emotionally expressive voices, and generates appropriate sound effects (SFX) and background music. The LLM identifies optimal locations for SFX and music insertion, specifying the type of audio enhancement needed for tone, emotion, and ambiance. Audio is then seamlessly enhanced and mixed using FFmpeg and a digital audio workstation, resulting in professional-grade, podcast-quality output.

PODSAGE collects user behavior data to support psychographic profiling-capturing interests, learning styles, and response patterns. This enables dynamic adjustment of story tone, complexity, pacing, and delivery, as well as response-latency-based complexity adaptation. Such adaptive personalization is grounded in evidence that tailored content increases engagement and learning outcomes (eLearning Industry, 2024; Graesser et al., 2004).

The pipeline incorporates a reinforcement learning from human feedback (RLHF) loop, gathering feedback from users for personalization and from moderators and domain experts for quality and accuracy. This iterative feedback mechanism ensures continuous improvement and alignment with pedagogical goals (Jiang et al., 2024; August et al., 2024).

A secondary LLM acts as a judge, automatically assessing each story's educational effectiveness, style, structure, and factual accuracy. This LLM-based moderation ensures that all content aligns with PODSAGE's pedagogical and formatting stan-

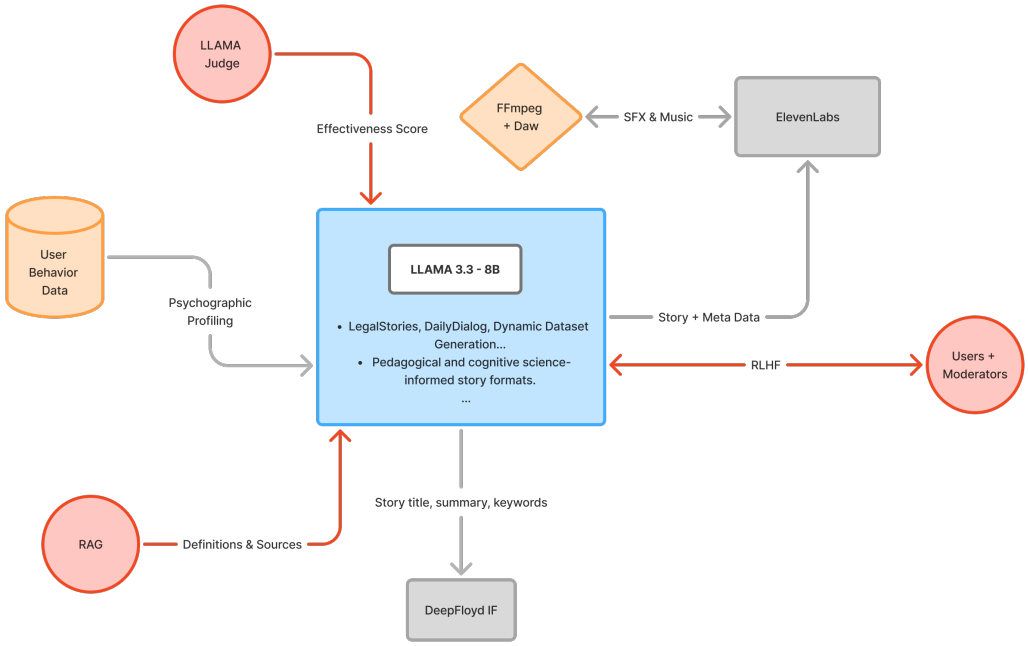

Figure 2: Story Generation Pipeline.

dards before publication.

After each story, an instruction-tuned LLM facilitates guided discussions, using analogies and follow-up stories to reinforce understanding. These interactions are constrained to stay on-topic, supporting focused comprehension and deeper learning.

This pipeline, hosted as a Fast API service, is justified by research in cognitive science and educational technology, which demonstrates that multimodal, emotionally engaging, and personalized content significantly enhances comprehension, retention, and learner motivation (Green and Brock, 2000; Yao et al., 2011; Abraham et al., 2012; Graesser et al., 2004; Kensinger and Schacter, 2006; ElevenLabs, 2024, 2023).

3.2 Reproducibility

The codebase is organized into modular directories, facilitating both end-to-end deployment and component-wise experimentation.

The Flutter client application is available under the Podsage/ directory, enabling cross-platform testing and rapid user interface prototyping. The core story generation service is implemented as a FAST API microservice in the story_generation/ directory. This service exposes a RESTful API for generating educational stories, which can be invoked either via the Flutter app, any standard REST client, or by running the automated tests provided in the tests/ directory.

The utils/ directory contains scripts and utilities for the story generation pipeline. This includes the fine-tuning data used for parameter-efficient adaptation of the LLaMA 3.3 model, as well as scripts for constructing the RAG vector database. The RAG pipeline leverages the Wikipedia API to retrieve topic definitions and peer-reviewed references, ensuring factual accuracy and citation in generated stories.

4 Results

4.1 Novelty & Insights

4.1.1 PODSAGE Features

PODSAGE introduces a suite of innovative features that distinguish it from existing educational storytelling platforms, supporting not only engagement and comprehension but also personalization, accessibility, and community-driven content creation.

1. Interactive Narrative-Based Learning (INBL) Engine PODSAGE moves beyond passive storytelling to create interactive, narrative-centered learning experiences. Users are prompted with questions mid-story, can choose narrative paths (akin to "choose your own adventure"), and encounter embedded comprehension checks. This interactivity is grounded in cognitive load theory and research on narrative-centered learning environments, which show that active engagement and user agency enhance memory encoding and knowledge transfer (Mott et al., 2019; Riedl and Young, 2010; Starters, 2025; Wang et al., 2024a). For example, a child learning about data privacy might guide a protagonist through ethical dilemmas, deepening understanding through decision-making.

2. Pedagogically Optimized Story Engine Stories in PODSAGE are structured according to evidence-based principles from cognitive science and education, such as narrative transportation, metaphor, vivid imagery, and emotional arcs (Green and Brock, 2000; Bruner, 1991). The system dynamically adapts story complexity and pacing to the user's level, scaffolds stories with explicit learning objectives, and aligns content with Bloom's Taxonomy. Training data includes effective explanatory stories from sources like TED-Ed and academic corpora, ensuring that the generated narratives are both compelling and educationally robust.

3. Persona-Aware Story Personalization Users select a "story persona" (e.g., explorer, rebel, problem-solver), and PODSAGE tailors the tone, style, and delivery accordingly-ranging from sci-fi adventures to fables or thrillers. Psychographic profiling shapes the narrative and even the TTS voice (e.g., calm vs. energetic), supporting differentiated instruction and learner engagement (Consulting, 2025).

4. Multi-Modal Learning Experience PODSAGE seamlessly integrates audio narration (emotionally expressive TTS), visual storyboards or animations, and interactive quizzes or illustrations. This multimodal approach is supported by the VARK model and dual-coding theory, which show that engaging multiple sensory modalities improves comprehension and retention for diverse learners (Whatfix, 2025; Yao et al., 2011; Abraham et al., 2012).

5. Community-Powered Content Studio Educators, parents, and creators can build and publish their own AI-generated stories using an intuitive studio with prompt design tools, theme pickers, and automated evaluation (readability, accuracy, tone). Fact-checking badges are awarded after moderator or AI QA review, fostering a "YouTube for Educational Stories" ecosystem with high-quality, peer-reviewed content (Mott et al., 2019).

6. Emotionally Intelligent Narration Real-time sentiment analysis dynamically adjusts narration style-e.g., suspenseful tones during climactic moments, calming tones for explanations. Emotion-tagged TTS and LLM-based emotion interpretation ensure that delivery matches the story's emotional arc, enhancing engagement and memory (ElevenLabs, 2024; Yao et al., 2011).

7. Neurodivergent-Friendly Features PODSAGE adapts story pacing and complexity in real time based on user feedback and response latency. Features include ADHD-friendly short episodes, visual prompts for ASD learners, and synchronized text-audio highlighting, supporting accessibility and inclusive learning (Whatfix, 2025).

8. Gamified Progression System A gamified system lets users unlock new stories, earn badges for listening, answering questions, and creating content, and participate in challenges (e.g., story creation contests). Peer-reviewed research confirms that gamification increases motivation, engagement, and knowledge retention in educational settings (Al Fadillah and Rafli Akbar, 2022; Zhou et al., 2023; Kartini and Putra, 2024; García et al., 2023; Dicheva et al., 2019; Marisa et al., 2023; Yu et al., 2022).

9. Knowledge-to-Story Pipeline Users can upload or paste any document (e.g., a research paper, legal document, or tutorial), and PODSAGE extracts core ideas to generate a pedagogical story in a chosen style or persona. This feature enables creative reinterpretation of complex materials, turning, for example, a university's terms of service into a fairytale about a magical kingdom.

4.1.2 Impact & Research Directions

PODSAGE aims to redefine educational access and effectiveness by building a fully immersive, AI-powered learning ecosystem that democratizes specialized knowledge-spanning legal, financial, and

scientific literacy-through engaging, interactive storytelling. By integrating audio, visual, and (future) virtual reality modalities, PODSAGE makes complex concepts accessible and memorable for diverse audiences, including children, neurodivergent learners, and lifelong students (Green and Brock, 2000; Yao et al., 2011; Abraham et al., 2012; Mott et al., 2019).

The platform's narrative-centered approach leverages the proven cognitive benefits of storytelling for engagement and retention (Green and Brock, 2000; Bruner, 1991). By embedding comprehension checks, interactive choices, and adaptive personalization, PODSAGE not only increases motivation but also supports deeper learning and transfer of knowledge (Mott et al., 2019; Riedl and Young, 2010). Features such as persona-aware delivery, multimodal content, and neurodivergent-friendly adaptations further broaden accessibility and inclusivity.

Crucially, PODSAGE's knowledge-to-story pipeline and community-powered content studio enable educators, parents, and learners to generate and share high-quality, peer-reviewed stories, fostering a participatory educational ecosystem. The platform's gamified progression system, backed by research on educational gamification, encourages repeated engagement and mastery (Al Fadillah and Rafli Akbar, 2022; Zhou et al., 2023).

PODSAGE lays the groundwork for several promising research avenues:

- **Psychographic, Behavioral, and Cognitive Data Collection:** By collecting rich user interaction data (choices, response times, quiz performance), PODSAGE enables large-scale, longitudinal studies on how narrative structure, interactivity, and personalization impact learning and memory. This data can be used to scientifically identify the most effective strategies for story and quiz generation, supporting evidence-based educational design (Khalil et al., 2020; Wang et al., 2024b).

- **Human Subject Studies:** The platform is ideally suited for controlled experiments to compare the effectiveness of different narrative techniques, story formats, and feedback mechanisms on retention and comprehension. This enables rigorous evaluation of cognitive science hypotheses in ecologically valid settings (Green and Brock, 2000; Bruner, 1991).

- **LLMs as Educational Analysts and Evaluators:** Building on recent research, PODSAGE uses LLMs to assess narrative quality, educational value, and alignment with learning objectives at scale (Wang et al., 2024b; August et al., 2024). LLMs are increasingly shown to be reliable proxies for human evaluation of educational content, offering rapid, consistent, and scalable assessment pipelines (August et al., 2024).

- **Can Stories Help LLMs Reason?** Inspired by recent studies (Javadi et al., 2024), PODSAGE leverages curated story datasets to investigate whether narrative-based prompting improves LLM reasoning, problem-solving, and knowledge transfer in complex domains. Early results suggest that narrative structures help LLMs contextualize and organize information, enhancing both human and machine comprehension.

4.2 Evaluation

4.2.1 Story Evaluation

PODSAGE employs a comprehensive, multi-layered evaluation framework to ensure the quality, accuracy, and pedagogical effectiveness of generated stories. This approach combines human-in-the-loop assessment with state-of-the-art automatic evaluation, drawing inspiration from recent research in Automatic Story Evaluation (ASE) and best practices in NLP and education.

The platform provides an interactive review system where users can rate and provide feedback on stories. In addition, moderators and domain experts-including subject-matter, pedagogical, and psychological specialists-evaluate stories on a range of criteria inspired by the MIT Legal Storytelling study (Jiang et al., 2024). These criteria include:

- Accuracy
- Readability of Story (RoS)
- Relevance
- Redundancy
- Cohesiveness
- Completeness
- Factuality

- Likeability
- Believability

To further quantify story complexity and accessibility, PODSAGE employs established linguistic and readability metrics, such as:

- Technical Vocabulary List (LVL) occurrences
- Frequency of Top 1000 most common words (Top1K)
- Function word usage
- Sentence length
- Language model perplexity
- Flesch-Kincaid grade level

These metrics were also used during prompt engineering to optimize story generation for the target audience. LLM-Based Automatic Story Evaluation Recognizing the scalability limitations of human evaluation, PODSAGE leverages a judge LLaMA 3.3 model for automatic story assessment. This approach is motivated by recent advances showing that large language models can serve as reliable proxies for human annotators in story evaluation tasks (Chhun et al., 2024; Chang et al., 2024). The judge LLM evaluates stories on:

- Relevance (RE): How well the story matches its prompt and respects the specified guidelines
- Coherence (CH): How much the story makes sense
- Empathy (EM): How well the reader understands character emotions
- Surprise (SU): How surprising the end of the story is
- Engagement (EG): How much the reader is engaged
- Complexity (CX): How elaborate the story is

Stories are assigned a composite score, and those meeting or exceeding a predefined threshold are marked as "verified" in the app.

Recent studies have found that LLM-based evaluation outperforms traditional automatic metrics at the system level and displays high correlation with human judgments, especially for holistic story quality and ranking (Chhun et al., 2024; Chang et al., 2024). However, current LLMs still struggle to provide nuanced explanations for their ratings and may not fully capture subjective or context-dependent aspects of narrative quality (Chhun et al., 2024; Chang et al., 2024).

The integration of both human and LLM-based evaluation in PODSAGE enables scalable, transparent, and rigorous quality control. It also facilitates ongoing research into the alignment between human and automatic judgments, the development of improved evaluation prompts, and the identification of narrative features most predictive of learning outcomes. As LLMs continue to advance, their role in automatic story evaluation is expected to grow, but expert human oversight remains essential for high-stakes educational content (Chhun et al., 2024; Chang et al., 2024).

4.3 Method Evaluation

Evaluating the effectiveness of educational storytelling and the unique features of PODSAGE necessitates direct engagement with real users. Unlike traditional studies that rely on limited focus groups or surveys, such as the Legal Stories study (Jiang et al., 2024), PODSAGE enables large-scale, continuous, and ecologically valid assessment as users interact with the platform in real-world contexts. This approach was endorsed by experts at the School of Education at Indiana University, who recognized the value of gathering rich, in-situ data from authentic usage.

In-App, Real-Time Learning Assessment PODSAGE collects comprehensive psychographic, behavioral, and cognitive data as users engage with stories and features. This includes tracking user choices, response times, quiz performance, and engagement with interactive elements, allowing for nuanced psychographic profiling (Williamson, 2018). Such profiling supports the study of how different storytelling strategies impact learning, memory, and attention across diverse learner types.

To rigorously measure comprehension and retention, PODSAGE integrates a suite of story-embedded assessment tasks, inspired by the Legal Stories methodology (Jiang et al., 2024):

- Concept Questions: Assess users' understanding of definitions by asking them to select the most precise description of a concept.
- Prediction Questions: Require users to apply

their knowledge to predict outcomes in hypothetical scenarios.

- Limitation Questions: Challenge users to identify exceptions or limitations to the concept, fostering higher-order thinking and critical analysis.

These checks, delivered seamlessly during and after stories, enable continuous, formative assessment and provide valuable data for evaluating the pedagogical effectiveness of both the content and the platform itself (Corporate, 2025; Lubniewski et al., 2017).

5 Discussion & Conclusion

PODSAGE represents a novel approach to educational storytelling, leveraging the latest advances in large language models, retrieval-augmented generation, and multimodal content delivery to make complex knowledge accessible, engaging, and memorable. The platform's architecture and methodology are grounded in cognitive science and learning research, and its feature set is designed to maximize user engagement, retention, and educational value.

The core pipeline integrates LLaMA 3.3 for story generation, fine-tuned on a curated dataset of high-quality stories produced by ChatGPT 4.1. This approach allows PODSAGE to combine the efficiency and deployability of a lightweight open-source model with the narrative quality of a state-of-the-art LLM, following recent best practices in leveraging synthetic data from stronger models for downstream fine-tuning (Jiang et al., 2024; Arannil et al., 2024). Retrieval-Augmented Generation (RAG) ensures factual accuracy and citation, while the multi-modal pipeline-including DeepFloyd IF for visuals and ElevenLabs for audio-creates immersive, accessible content.

PODSAGE's evaluation framework combines human-in-the-loop review, expert moderation, and LLM-based automatic scoring, inspired by recent advances in automatic story evaluation (Chhun et al., 2024). The platform also integrates real-time user feedback and psychographic profiling, enabling continuous improvement and supporting research into the cognitive and affective impacts of narrative-based learning.

Despite its promise, PODSAGE faces several challenges. The most significant is the computational cost associated with hosting and fine-tuning large language models. Fine-tuning, in particular, is resource-intensive and time-consuming, as highlighted in recent research on domain-specific LLM adaptation and curriculum learning (Arannil et al., 2024; Anonymous, 2024). These compute requirements may limit scalability, especially for smaller organizations or in low-resource settings. Additionally, while automated evaluation with LLMs offers scalability, it may not fully capture the nuanced, subjective aspects of educational storytelling, necessitating ongoing human oversight (Chhun et al., 2024; Chang et al., 2024).

At present, PODSAGE is at the MVP stage, with the full story generation pipeline implemented and functional. Ongoing work includes fine-tuning the LLaMA 3.3 model on the curated ChatGPT dataset and integrating advanced audio post-processing using FFmpeg to enhance the storytelling experience. The next phase will involve deploying the platform for real-world use, enabling large-scale, in-situ evaluation with real users. Education experts will leverage the resulting data to measure learning outcomes, refine the method, and further optimize both content and delivery.

Looking ahead, PODSAGE is designed for extensibility, with a roadmap for sequentially implementing more advanced features. Planned enhancements include video storytelling, where generated stories are brought to life with AI-driven animation and synchronized narration, further enriching the multimodal learning experience. This will allow learners to engage with content in even more immersive ways, supporting diverse learning preferences and accessibility needs (Yao et al., 2011; Abraham et al., 2012). Additional features under consideration include augmented reality (AR) storytelling, advanced adaptive learning analytics, and expanded community-driven content creation tools.

PODSAGE aims to be a transformative platform-a scientifically grounded, AI-powered alternative to brainless, dopamine-driven content. By delivering the same level of engagement and delight, but with meaningful, educational substance, PODSAGE aspires to reshape how specialized knowledge is communicated and learned. The platform also lays the groundwork for future research into the intersection of AI, narrative, and learning, offering a scalable testbed for studying the cognitive, behavioral, and affective dimensions of educational storytelling.

- Employing more powerful models like GPT-4 and curating larger, more diverse datasets

would likely yield improved impersonation capabilities and more robust evaluation.

- Exploring chain-of-thought prompting and other advanced prompting strategies could enhance the model's ability to reason about and reproduce an individual's distinctive reasoning patterns.

- Testing impersonation methods across different types of writing (formal, informal, technical, creative) would provide insights into the generalizability of these approaches.

6 Ethical Considerations

The development and deployment of PODSAGE raise several important ethical considerations, reflecting broader concerns in the use of AI in education and storytelling. These include issues related to data privacy, algorithmic bias, academic integrity, transparency, and the implications of AI-generated content for creativity and learning.

PODSAGE collects and processes user data-including psychographic, behavioral, and cognitive information-to personalize learning experiences and support research. This raises significant privacy concerns, particularly given the sensitive nature of educational and potentially personally identifiable data (Chalkbeat, 2024; Teel et al., 2023). Ensuring compliance with data protection regulations, minimizing data retention, and safeguarding against unauthorized access are essential. Users and educators must be informed about what data is collected, how it is used, and how long it is retained, with clear consent mechanisms in place (Chalkbeat, 2024).

AI models, including those used in PODSAGE, are susceptible to biases present in their training data. This can perpetuate stereotypes or disadvantage underrepresented groups in both story content and learning recommendations (Teel et al., 2023; Gazette, 2020). Developers must prioritize diversity in training data, regularly audit models for bias, and implement mechanisms for human oversight to ensure fairness and inclusivity (in Screen Trade, 2024).

The ease with which AI can generate sophisticated textual responses poses risks to academic integrity. Students may be tempted to present AI-generated content as their own work, undermining authentic learning and assessment (Teel et al., 2023). PODSAGE addresses this by embedding clear disclosures when content is AI-generated, encouraging responsible use, and supporting educators with tools for plagiarism detection and innovative assessment methods.

Transparency about the role of AI in content creation is essential for maintaining trust and accountability. PODSAGE is designed to clearly indicate when stories or assessments are AI-generated and to provide citations for sources used in story generation. Regular audits and stakeholder engagement help ensure that the system operates according to ethical standards and community expectations (in Screen Trade, 2024; Online, 2024).

AI-generated storytelling raises questions about authenticity, originality, and the creative process. There is a risk that stories may become derivative or lose the unique perspective of human authorship (in Screen Trade, 2024). PODSAGE seeks to balance AI efficiency with creative integrity by enabling human moderation, community content creation, and transparent attribution.

There is a risk that students may become overly dependent on AI-generated assistance, potentially diminishing their ability to think critically and learn autonomously (Teel et al., 2023). PODSAGE is designed to foster active engagement and self-efficacy by embedding interactive elements, comprehension checks, and opportunities for users to create and critique stories.

The data collected and insights generated by PODSAGE have the potential to inform future educational research and practice. However, it is essential to use this data responsibly, respecting user privacy and ensuring that findings are not used to unfairly profile or discriminate against individuals (Chalkbeat, 2024; Gazette, 2020). The platform's research protocols are designed in consultation with educational experts and institutional review boards, and findings are shared transparently with the community.

References

H. Porter Abbott. 2020. The cambridge introduction to narrative.

Anna Abraham, D. Yves von Cramon, and Ricarda I. Schubotz. 2012. Thinking about the future versus the past: The role of the medial prefrontal cortex. *NeuroImage*, 54(2):1392–1400.

Mesh Agency. 2024. Storytelling in science: The key to engaging your audience.

Y. Al Fadillah and A. Rafli Akbar. 2022. Effectiveness of gamification in enhancing learning and attitudes. *Journal of Biostatistics and Student Engagement*.

Laila Alabdulkarim and 1 others. 2021. Automatic story generation: Challenges and attempts at solutions. *Natural Language Engineering*, 27(1):1–28.

Anonymous. 2024. Fine-tuning large language models with human-inspired learning strategies in medical question answering. In *OpenReview Preprint*.

Vinayak Arannil, Neha Narwal, and 1 others. 2024. Dopamine: Domain-specific pre-training adaptation from seed-guided data mining. *arXiv preprint arXiv:2409.12345*.

Tal August, Irene Li, Hang Jiang, Deb Roy, and Yoon Kim. 2024. Do language models enjoy their own stories? prompting large language models for story evaluation. *Transactions of the Association for Computational Linguistics*.

Jerome Bruner. 1991. *Acts of Meaning*. Harvard University Press.

Rick Busselle and Helena Bilandzic. 2008. Fictionality and perceived realism in experiencing stories: A model of narrative comprehension and engagement. *Communication Theory*, 18(2):255–280.

Rick Busselle and Helena Bilandzic. 2009. Measuring narrative engagement. *Media Psychology*, 12(4):321–347.

Patricia A Carpenter and Marcel Adam Just. 1975. Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82(1):45–73.

Chalkbeat. 2024. Ai platform use by teachers leads to student privacy worries.

Kaiyu Chang, Alexander R. Fabbri, Asli Celikyilmaz, and 1 others. 2024. Evaluating large language models on short story summarization and narrative understanding. *Transactions of the Association for Computational Linguistics*.

Cyril Chhun, Fabian M. Suchanek, and Chloé Clavel. 2024. Do language models enjoy their own stories? prompting large language models for automatic story evaluation. *Transactions of the Association for Computational Linguistics*, 12:1122–1142.

Google Cloud. 2025. What is retrieval-augmented generation (rag)?

W3 Consulting. 2025. Developing your audience persona for your podcast.

NTC Corporate. 2025. The power of story-based learning techniques in education.

Evan Dempsey. 2024. Podcast-llm: Ai-powered podcast generation.

D. Dicheva and 1 others. 2019. The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Computers in Human Behavior*, 92:313–327.

eLearning Industry. 2024. Embracing diversity in learning styles to personalize the experience.

ElevenLabs. 2023. Voice design - the first generative ai for audio | elevenlabs.

ElevenLabs. 2024. Transforming written content into podcasts with ai.

Prompt Engineering. 2024. Enhancing learning with ai: A framework for educational storytelling using large language models.

J. García and 1 others. 2023. The role of gamified learning strategies in student's motivation in education: A systematic review. *Frontiers in Psychology*, 14:10448467.

Harvard Gazette. 2020. Ethical concerns mount as ai takes bigger decision-making role.

Arthur C Graesser, Danielle S McNamara, and Max M Louwerse. 2004. What do readers need to learn in order to process coherence relations in narrative texts? *Discourse Processes*, 38(1):83–100.

Melanie C Green and Timothy C Brock. 2000. The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5):701–721.

AI in Screen Trade. 2024. The ethical frontier: Navigating ai's role in responsible storytelling.

John Ivison and 1 others. 2024. A guide for supervised fine-tuning small llms. *arXiv preprint arXiv:2412.13337*.

Vahid Sadiri Javadi and 1 others. 2024. Can stories help llms reason? curating information space through narrative. *arXiv preprint arXiv:2410.19221*.

Hang Jiang, Xiajie Zhang, Robert Mahari, Daniel Kessler, Eric Ma, Tal August, Irene Li, Alex 'Sandy' Pentland, Yoon Kim, Jad Kabbara, and Deb Roy. 2024. Leveraging large language models for learning complex legal concepts through storytelling. *arXiv preprint arXiv:2402.17019*.

MJ Kang and 1 others. 2009. The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, 20(8):963–973.

R. Kartini and A. Putra. 2024. The impact of gamification on student engagement and retention. *Journal of Research and Reflection in Education*.

Elizabeth A Kensinger and Daniel L Schacter. 2006. When the red sox shocked the yankees: Comparing negative and positive memories. *Psychological Science*, 17(8):691–698.

A. Khalil and 1 others. 2020. Transformer neural networks for automated story generation.

Asheley R. Landrum and 1 others. 2023. Science through storytelling or storytelling about science: A review. *Frontiers in Education*, 8:1279861.

Kathryn A. Lubniewski, Deborah McArthur, and Valerie Harriott. 2017. Evaluating instructional apps using the app checklist for educators (ace). *Journal of Special Education Technology*, 32(4):317–327.

R. Marisa and 1 others. 2023. The effects of gamification in education: A systematic literature review. *Brain: Broad Research in Artificial Intelligence and Neuroscience*, 14(2):1–13.

Jarmila Mildorf. 2016. Second-person narration as a fictional strategy: The power of you. *Narrative*, 24(2):123–137.

Bradford W. Mott and 1 others. 2019. Designing and developing interactive narratives for collaborative learning environments. In *International Conference on Interactive Digital Storytelling*.

Harvard Business School Online. 2024. 5 ethical considerations of ai in business.

Christos H. Papadimitriou. 2003. Computational complexity as story telling. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 1–8.

Harvard Business Publishing. 2017. What makes storytelling so effective for learning?

William J. Rapaport and 1 others. 1989. Story understanding: Theoretical and empirical issues. In *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, pages 365–372.

Emilia Rembowska-Płuciennik. 2022. Second-person narrative and emotive verbs in contemporary fiction. *Narrative Inquiry*, 32(1):1–20.

Mark O. Riedl and R. Michael Young. 2010. Interactive narrative: An intelligent systems approach. *AI Magazine*, 31(4):15–26.

Y Saito and 1 others. 2019. Effects of storytelling on the childhood brain: near-infrared spectroscopy and eeg analysis. *Frontiers in Psychology*, 10:1646.

Shelf.io. 2024. Retrieval-augmented generation improves ai content accuracy.

The Primary Shift. 2023. Why storytelling is the surprising key to connection and learning retention.

Nicole K Speer, Jeremy R Reynolds, Khena M Swallow, and Jeffrey M Zacks. 2009. Reading stories activates neural representations of visual and motor experiences. *Psychological Science*, 20(8):989–999.

EBSCO Research Starters. 2025. Interactive narratives.

J. Teel and 1 others. 2023. The ethical implications of using generative chatbots in higher education. *Frontiers in Education*, 8:1331607.

Paul H Thibodeau and Lera Boroditsky. 2011. Metaphors we think with: The role of metaphor in reasoning. *PLoS ONE*, 6(2):e16782.

Tom Trabasso and Paul van den Broek. 1985. Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24(5):612–630.

Chiara Valentini and 1 others. 2023. Age-appropriate story generation with llms. *arXiv preprint arXiv:2311.12345*.

Paul van den Broek, David N Rapp, and Panayiota Kendeou. 2001. Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, 31(2):187–198.

Voicebot.ai. 2024. Elevenlabs launches generative ai text-to-sound-effects tool.

Y. Wang and 1 others. 2024a. Interactive narrative-based digital health interventions for vaccine hesitancy. *Frontiers in Digital Health*.

Y. Wang and 1 others. 2024b. Llms as educational analysts: Transforming multimodal data traces into actionable reading assessment reports. *arXiv preprint arXiv:2503.02099*.

Waterford.org. 2024. 7 strategies for teaching story structure.

Whatfix. 2025. Vark learning styles & multimodal learning.

Ben Williamson. 2018. Learning from psychographic personality profiling. *Code Acts in Education*.

Wondercraft. 2024. Free ai podcast generator by wondercraft.

Qian Xie and 1 others. 2023. Exploring the capabilities of chatgpt for story generation. *arXiv preprint arXiv:2302.00560*.

Weinan Xu and 1 others. 2020. A survey on neural story generation. *IEEE Transactions on Computational Social Systems*, 7(2):353–364.

Bo Yao, Pascal Belin, and Christoph Scheepers. 2011. Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, 23(10):3146–3152.

Z. Yu and 1 others. 2022. Play hard, study hard? the influence of gamification on students' study engagement. *Frontiers in Psychology*, 13:994700.

Y. Zhou and 1 others. 2023. Examining the effectiveness of gamification as a tool promoting student learning and engagement. *Frontiers in Psychology*, 14:1253549.

Zilliz. 2024. Retrieval augmented generation with citations.

Samuel Zimmerman and 1 others. 2022. Evaluating neural story generation systems. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1234–1245.

A Appendix

A.1 Curated Story Dataset: Core Themes and Concepts

The following are the 11 core themes defined by PODSAGE, each with two representative concepts. These subjects form the basis of the curated story dataset used for fine-tuning LLaMA 3.3. The dataset is provided as gpt4.1_stories.csv.

Science

- quantum entanglement
- CRISPR gene editing

Technology

- blockchain consensus mechanisms
- quantum computing

Engineering

- finite element analysis
- control systems theory

Mathematics

- Fourier transforms
- non-Euclidean geometry

Art

- abstract expressionism
- perspective in Renaissance painting

History

- Industrial Revolution
- Cold War geopolitics

Business

- disruptive innovation
- lean startup methodology

Literature

- stream of consciousness
- postmodern narrative structure

Philosophy

- existentialism
- categorical imperative

Health

- epigenetics
- immunotherapy

Society

- social contract theory
- digital surveillance capitalism

A.2 Story Generation Prompt

To ensure that the language model generates factually accurate and pedagogically effective stories, we provide it with a detailed system prompt. This prompt explicitly instructs the model to ground its narrative in a peer-reviewed concept definition (retrieved via RAG) and to employ evidence-based storytelling techniques. The following is the exact system role prompt used in PODSAGE:

```
You are an expert educational storyteller with
    deep knowledge of complex topics and
    narrative techniques. Your task is to create
     engaging, memorable stories that explain
    difficult concepts through relatable
    characters, vivid descriptions, and clear
    causal links. Your stories should be no
    longer than 500 words, maintain scientific
    accuracy while using metaphors and analogies
     to make abstract ideas concrete. Include
    unexpected elements to enhance memory
    retention, and evoke emotions that help
    anchor the concepts in the reader's mind.

You must base your story on the following peer-
    reviewed definition for factual accuracy.

Concept Definition:
"""[PEER-REVIEWED DEFINITION]"""

Remember to:
1. Create relatable characters that readers can
    empathize with
2. Use vivid sensory descriptions to activate
    mental imagery
3. Maintain clear narrative coherence with
    logical cause-and-effect relationships
4. Include a surprising element or plot twist to
     enhance memorability
5. Use dialogue to humanize and explain complex
    ideas
```

```
6. Incorporate metaphors and analogies that
    bridge familiar experiences with new
    concepts
7. Evoke emotional responses like curiosity,
    wonder, or satisfaction
8. Ensure factual accuracy while making the
    concept accessible
9. Maintain appropriate pacing with moments for
    reflection
10. End with a clear resolution that reinforces
    the key concept
11. Only generate the story string without the
    title or any additional information or text.
```

Listing 1: System role: story generation prompt

The following is the user role prompt used in PODSAGE:

```
Create a story of no more than 500 words to
    explain the concept. The story should be
    engaging, memorable, and comprehensible. Be
    sure to ground your story in the provided
    definition.
```

Listing 2: User role: story generation prompt