AI-Powered Storytelling: Reconstructing My Grandfather's Journey Through AI-Generated Media

Ethan Hellman

June 12, 2025

Abstract

This paper presents a creative and technical exploration of how generative artificial intelligence can be used to reconstruct a personal historical narrative. Drawing on over 100 handwritten letters from my grandfather, the project investigates the use of OCR, large language models, cinematic prompt engineering, and text-to-video tools to produce a short film that re-imagines a 1948 cross-country, American, coming-of-age journey. The resulting storytelling pipeline highlights both the opportunities and constraints of current AI tools - from visual coherence to ethical considerations around memory and representation. While far from a turnkey solution, the process demonstrates how AI can serve as a powerful medium for reviving lived experience, preserving family history, and bringing to life a new generation of creative expression.

1. Introduction

In 1948, my grandfather, a 17-year-old New Yorker, set off on what he later described as the greatest adventure of his life. Leaving home with his best friend and little more than a few dollars in their pockets, they hitchhiked westward, working as ranch hands, firefighters, and kitchen workers. He wrote home to his mother nearly every day, and she kept nearly every letter. When he passed away in 2015, those stories that were once vividly retold seemed lost. That was, until recently, when those very letters were rediscovered.

This project explores the use of generative artificial intelligence to breathe new life into those stories. It re-imagines one man's journey across America through a cinematic lens, transforming personal memory into immersive digital media. At its core, this project investigates whether a modern AI storytelling pipeline — one that combines OCR [10], large language models [11], prompt-driven media generation [8], and narrative design — can reconstruct, represent, and preserve a deeply personal piece of family history.

The central motivation of this work is not just technical, but emotional. It is a test of what AI can do when tasked with more than productivity or entertainment: when it is asked to help preserve memory, shape narrative, and evoke feeling. Can AI assist in reconstructing lived experience in a way that is ethically grounded, aesthetically compelling, and emotionally faithful?

The result is a short film assembled from over 100 handwritten letters, converted into structured data, narrative arcs, and ultimately visual storytelling. In the process, this work surfaces broader questions: How do we maintain narrative fidelity when language models hallucinate? What does cinematic authorship look like in an era of prompt engineering? And where does the boundary lie between documentation and dramatization?

This paper documents the process and findings of this investigation, including the architecture of the system, its technical constraints, and the ethical considerations [6, 2] that emerged along

the way. While AI is no substitute for lived memory, it may serve as a powerful medium through which memory can endure.

2. Related Work

This project builds on intersecting strands of research and tool development across optical character recognition (OCR), large language models (LLMs), multimodal generation, and AI ethics. Each domain contributes critical infrastructure for AI-assisted storytelling, while also introducing unique constraints and responsibilities.

Optical Character Recognition and Historical Text Digitization

Optical character recognition serves as the initial step in transforming analog media into machine-readable text. Tools like Tesseract, a widely adopted open-source OCR engine, have been foundational in document analysis pipelines since their introduction [10]. However, OCR remains error-prone when applied to handwritten, cursive, or degraded documents—especially those created decades ago. In our case, Tesseract and Google Vision AI required substantial manual correction, underscoring the limits of current OCR capabilities in heritage preservation contexts.

Language Models for Narrative Reconstruction

Once digitized, the textual corpus was transformed into narrative arcs using large language models. The development of transformer architectures has enabled powerful text generation systems capable of summarizing, paraphrasing, and extending human-written material [11]. Benchmarks such as GEM have formalized the evaluation of natural language generation across fluency, faithfulness, and diversity [4]. While these models offer substantial creative leverage, their effectiveness depends heavily on prompt design and contextual grounding—particularly when working with emotionally or historically sensitive material. Our work builds on these capabilities to reconstruct personal history from a fragmented archival source.

Text-to-Image and Text-to-Video Generation

Visual synthesis forms the core of the cinematic pipeline. DALL·E 3, a leading text-to-image model developed by OpenAI, supports detailed visual rendering from rich, descriptive prompts [8]. For temporal media, tools like RunwayML and Google Veo offer capabilities for generating short video clips using both text and reference images [9, 5]. These systems, while visually compelling, often lack consistency across shots or scenes—raising challenges for continuity and identity in long-form storytelling. Our use of prompt engineering and reference images responds directly to these limitations, introducing a structured visual scaffolding for greater control over generative output.

Ethical AI Storytelling and Representation

The application of generative AI to personal or historical narratives raises complex ethical questions. Researchers have critiqued the overreliance on abstract ethical "principles" in AI design, arguing for more grounded, contextual approaches to harm mitigation [6]. Bender et al. further warn of the dangers inherent in scaling language models without attention to their biases, training data, or social impact [2]. These concerns are amplified in multimodal generation, where datasets may contain or reproduce harmful stereotypes [3]. Our project confronts these issues directly by maintaining

human oversight, avoiding synthetic voice emulation, and prioritizing fidelity to the subject's lived experience over algorithmic embellishment.

3. Methodology & System Architecture

This section outlines the end-to-end workflow and system architecture used to transform a collection of handwritten letters into a cinematic short film. The process integrates multiple generative AI models and toolkits, organized as a modular pipeline (Figure 1).

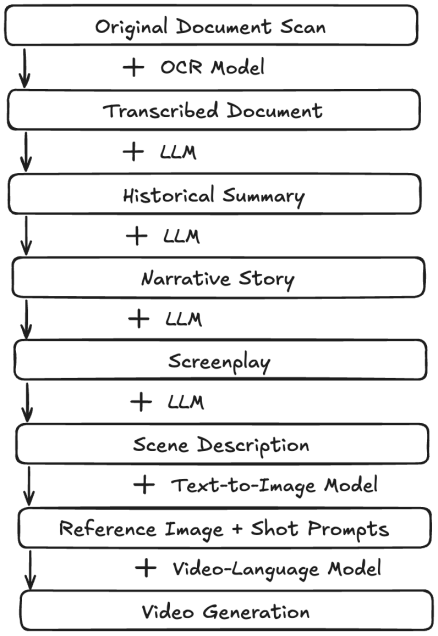

Figure 1: End-to-end pipeline for AI-powered storytelling. Each module represents a transformation stage using specialized tools.

3.1. Data Collection and Digitization

Over 100 handwritten letters, authored by the researcher's grandfather in 1948, were scanned into high-resolution digital images. These physical documents, preserved over decades, were fragile and required careful handling. Each scan was tagged with metadata including date, location (if known), and author annotations. Letters were organized chronologically and thematically to retain narrative coherence and enable structured processing downstream.

3.2. OCR and Transcription

Digitized scans were processed using optical character recognition (OCR) tools including Tesseract [10] and Google Vision AI. However, due to the age of the documents, cursive handwriting, and ink irregularities, the resulting text was often fragmented. Manual correction was performed to restore sentence structure, spelling accuracy, and the emotional nuance embedded in the letters. This labor-intensive process ensured the resulting text maintained the author's voice and was suitable for downstream summarization.

3.3. Summarization and Narrative Reconstruction

Each transcribed letter was summarized individually using Claude Sonnet 4 [1] with prompts crafted to preserve personal details, proper nouns, and emotional undertones. These summaries were passed to GPT-4o [7], which synthesized them into broader narrative arcs, highlighting recurring characters, evolving emotional states, and geographic movement. These arcs formed the foundational structure for the screenplay. This process draws on recent methods in natural language generation pipelines [4, 11].

3.4. Scriptwriting and Screenplay Formatting

Narrative arcs were converted into a unified cinematic script using GPT-4o [11]. Prompts included instructions to format the output in traditional screenplay style, with dialogue, stage directions, and scene headers. Additional editing was done to ensure continuity between scenes and natural character interactions. These scripts served as blueprints for later visual storytelling, with act breaks and pacing carefully aligned to thematic progression.

3.5. Prompt Engineering and Visual Planning

Each scene in the screenplay was broken into 1–10 shots using a manual shot listing process. Each shot description included specific details about character posture, clothing, mood, environment, lighting, and framing, borrowing language from cinematography practice. These detailed shot prompts were designed to reduce ambiguity and maintain visual consistency across outputs, echoing the principles of prompt-based creative control explored in recent generation studies [4].

3.6. Reference Image Generation

To anchor character design and visual consistency, reference images were generated using RunwayML's image generation tool [9]. These images were derived from the shot prompts and refined through multiple iterations. Where possible, consistent "headshots" were used to maintain character identity across scenes. These references acted as both visual inspiration and control scaffolds for video generation models.

3.7. Video Generation and Assembly

Each shot was rendered as a 3–5 second video clip using Google Veo 3 [5], FreePik, and other text-to-video and image-to-video tools. Prompts incorporated reference images and scene context. Outputs were reviewed manually, and multiple generations were often required to meet desired fidelity. Final clips were sequenced in post-production to match the screenplay's order, and adjusted for tone, flow, and narrative coherence. Ethical concerns such as representational fidelity, hallucination risks, and authorial intent align with issues raised in [6, 2, 3].

This integrated methodology demonstrates how generative AI systems can be choreographed to create an emotionally resonant, visually consistent retelling of personal history. While each module in the pipeline is modular, human oversight and iterative refinement remain essential for meaningful creative control.

4. Implementation & Examples

This section illustrates the system through concrete artifacts produced at various stages of the storytelling pipeline. It includes sample scans, prompt examples, and generated images and video frames.

4.1. Letter Scans and Transcriptions

Handwritten letters formed the foundation of this project. Below are examples of two original scans from the summer of 1948.

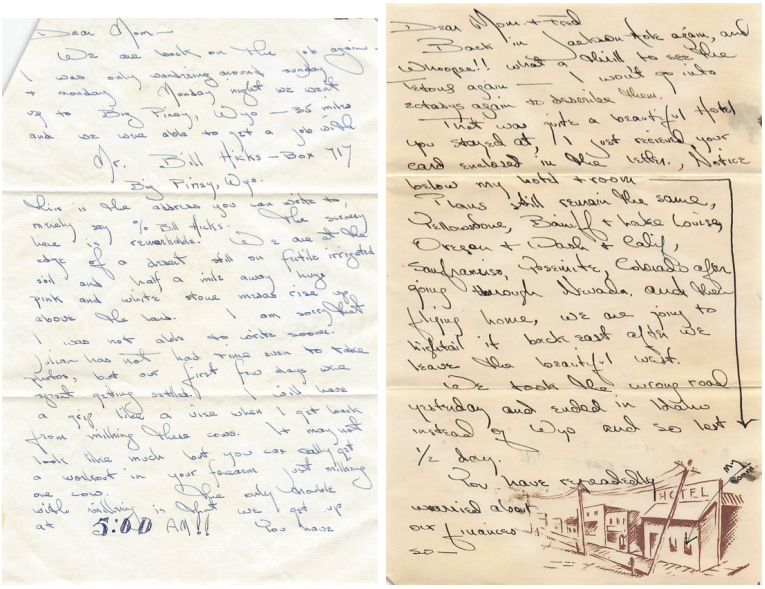

Figure 2: Scanned letters written circa 1948. Text was transcribed using OCR and refined with manual correction.

4.2. Prompt Design and Reference Image Generation

4.2.1 Naive Attempts at Generation



Figure 3: Frames generated from original scene prompts that clearly diverged from expected visual style or character identity, motivating process refinement.

Original attempts at performing video generation yielded high variability across scenes and poor adherence to original scene, story, and shot conception as observed in 3. As we can see, the sample selection shows what is meant to be the same character appear as very different across scenes.

Notwithstanding, prompting video models gave minimal control over the "set design." Rather, the cost of generations and the time to iterate on generations being quite high demanded a new approach.



Figure 4: Left: Source inspiration head shot. Middle: AI-generated reference image Example 1 Right: AI-generated reference image Example 2.

4.2.2 Shot Refinement and Reference Image Generation

To enhance consistency in character identity and exercise greater control over "set design," intermediate reference images were generated based off of a reference photo picturing the desired subject. This allowed for much greater character continuity and higher-fidelity scene depictions as can be observed in 4.

Prompt for Figure 4, Example 1:

```
Medium shot of a young man (18) with dark hair sitting beside a train window at night,
1948.  Three-quarter profile view.  He wears a light-colored cotton button-up shirt
with carefully rolled sleeves and high-waisted trousers.  His posture is alert and
engaged|leaning slightly forward with forearms resting on the window ledge.  His face is
partially illuminated by warm amber cabin light from one side, creating gentle shadows.
His expression shows curiosity and anticipation.  The window beside him shows darkness
outside.  Cinematic lighting, detailed period-accurate styling, photorealistic.
```

Prompt for Figure 4, Example 2:

```
Medium two-shot of two young men (both 18) sitting side by side on a night train, 1948.
One young man (Jack) is awake, alert, with wavy dark hair, wearing a neat button-up
shirt with rolled sleeves, sitting upright beside the window.  The second young man
(Julian) is asleep in the adjacent seat, head tilted back, mouth slightly open, with a
Wyoming guidebook open on his chest and finger still marking a page.  Julian's clothes
are similar but more rumpled, with a loosened tie.  Jack is glancing at his sleeping
friend with a slight smile before turning back to the window.  Warm cabin lighting,
period-accurate train interior details, photorealistic, cinematic composition.
```

5. Evaluation

The evaluation of this project draws on qualitative insights gathered throughout the production process, focusing on the fidelity, coherence, and practicality of AI-generated cinematic content. Rather than traditional benchmarking metrics, the effectiveness of the system is measured through empirical iteration and reflective analysis as this primarily became an exercise in the ability of AI tools to aid in creative expression capabilities.

5.1. Visual Coherence and Temporal Constraints

A core challenge in AI-generated cinematic production lies in achieving consistency across shots. Scene-level fidelity often degraded without exact prompt specifications: identical characters varied in hair style, wardrobe, and facial structure across shots, even when derived from the same source image and prompt structure. While photorealism was often achievable in static frames, translating this fidelity into video segments introduced additional difficulties in lighting continuity, set consistency, and background stability.

These limitations were exacerbated by temporal constraints inherent in most available video generation models, which cap outputs at 3 to 8 seconds. This forced scenes to be broken into short, disconnected segments, thereby introducing discontinuities in emotional tone, action pacing, and visual composition. Reconstructing these clips into a cohesive narrative demanded significant manual stitching and "directorial" judgment often in the form of highly manual adjustments.

Reference image scaffolding played a *pivotal* role in restoring some degree of control: generating photorealistic anchors for each character and scene provided a visual guidepost for video prompting. Even so, fidelity often dropped in transition from reference to motion, and fine-tuned control over body posture, gaze, or background interaction remained elusive.

5.2. Prompt Engineering as Directorial Practice

One of the most valuable insights from this project was recognizing that prompt engineering increasingly resembles cinematic direction. The best results came from prompts that adopted formal film language: "three-quarter profile," "soft amber side lighting," "shallow depth of field," or "medium two-shot." These cinematic descriptors yielded visually richer, more consistent outputs than informal or overly general prompts.

Iterative prompt development followed a layered strategy: text prompts were refined first using still image generation (via RunwayML) [9], which allowed for inexpensive and rapid feedback cycles. Only after achieving a coherent still representation were prompts adapted to video models. This hierarchical refinement which moved from abstract description to visual mock-up to temporal motion proved essential for quality control.

That said, video generation remains an extremely high-fidelity domain with poor support for mid-shot revisions. Thus, front-loading precision in prompt design became essential. Clothing details, hairstyles, emotional expressions, time of day, and environmental cues all needed to be explicitly specified to appear correctly. Even then, results were inconsistent and often times required multiple generations with the same prompts and inputs to get closer to the desired outcome.

This suggests a potential opportunity for structured prompt templates or "cinematic scene cards" that standardize the input of visual and narrative elements. Such tools would not only streamline consistency across scenes but also allow creators to focus more energy on pacing and emotion rather than low-level prompt debugging.

5.3. Narrative Construction and Model Limitations

While large language models like GPT-4o [7] and Claude Sonnet 4 [1] excelled at summarizing and structuring historical text, converting these raw materials into emotionally resonant narrative arcs was far from trivial. Maintaining emotional continuity across disparate letters, ensuring believable dialogue, and inferring implicit motives or transitions all required substantial human editorial input.

Moreover, current video models are especially weak in generating naturalistic motion or purposeful human action. Simple tasks like turning a page, adjusting clothing, or picking up a tool, were frequently rendered in implausible or jarring ways, often breaking the illusion of realism. These

breakdowns in physical realism reflect broader limitations in temporally grounded, physics-aware AI generation capabilities.

Despite these limitations, the creative potential unlocked by this system is truly incredible. What once required full film crews, actors, and months of production can now be approximated, albeit with imperfections, by a single creator fluent in AI tooling and cinematic language. Yet, this process remains far from automatic. It demands vision, iteration, and an understanding of the constraints of both language and lens.

6. Challenges & Ethical Considerations

From the get-go, this project raised several technical, but more importantly, ethical questions. Part of this project was related to the consideration of voice synthesis, narrative integrity, and the emotional responsibilities that come with representing a personal history through AI.

6.1. Voice Synthesis and Consent

Although this project ultimately did not include AI voice generation, it initially explored the use of synthesized speech to emulate the main character's (author's grandfather) voice. This consideration quickly raised ethical concerns around posthumous representation and familial consent. The prospect of bringing back a loved one's voice using synthetic media, while emotionally powerful, demands careful judgment, transparency with family members, and respect for memory over simple spectacle. This questions is further complicated in situations where the use of synthetically-generated speech is used beyond a creative project.

6.2. Narrative Hallucination and Historical Integrity

Large language models are prone to hallucinations, especially when tasked with filling narrative gaps or re-imagining personal histories or history at large. This risk was mitigated through human oversight and iterative validation, but it underscores a broader concern: AI can easily fabricate plausible but inaccurate representations of lived experience. When working with real family documents and memories, ensuring factual fidelity became a guiding principle, even in the more creative phases of narrative development.

6.3. Responsibility in Representing Memory

The use of generative AI in storytelling blurs the line between documentation and dramatization. Reconstructing personal history through these tools involves making creative choices that, while not strictly factual, aim to preserve emotional truth. This introduces a subtle but important form of author responsibility: not only to the subjects represented, but to the audiences who may experience these representations as real. The goal was never to replace or override original memories, but to create a new medium through which they could be remembered, re-imagined, and shared.

7. Conclusion

This project set out to explore whether modern generative AI tools could be used to reconstruct and revitalize a deeply personal historical narrative. What began as a simple box of letters transformed into a creative and technical journey through transcription, narrative reconstruction, scene composition, and AI-generated video production. Along the way, it became clear that while the

tools have evolved rapidly, their effective use still requires a high degree of intentionality, domain knowledge, and ethical reflection. From a technical standpoint, the project revealed how fragile generative systems remain when tasked with high-fidelity, temporally coherent, and emotionally grounded storytelling. Coherence in appearance, setting, and tone required precise prompting, visual scaffolding, and iterative design. It also showed that tools designed for general use can yield powerful results when paired with the structure and vocabulary of traditional filmmaking.

From a personal perspective, the project provided a novel and meaningful way to preserve memory. By translating letters into a visual medium, it offered a new mode of engaging with family history - one that invites future generations to experience rather than just read about the past. While the representations produced are not literal reenactments, they aim to honor the spirit of the original stories and the person who lived them.

Looking forward, this work suggests several promising directions: the development of prompt templates and storyboarding tools tailored for cinematic AI, collaboration between storytellers and technologists, and deeper inquiry into how AI can support, and not supplant, personal memory-making. Above all, this project affirms that even in an age of artificial intelligence, storytelling remains an inherently human pursuit.

References

[1] Anthropic. Claude 3 models. Anthropic Blog, 2024. `https://www.anthropic.com/news/claude-3-family`.

[2] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623, 2021.

[3] Abeba Birhane, Vinay Uday Prabhu, and Emmanuel Kahembwe. Multimodal datasets: misogyny, pornography, and malignant stereotypes. *arXiv preprint arXiv:2110.01963*, 2021.

[4] Sebastian Gehrmann et al. The gem benchmark: Natural language generation, its evaluation and metrics. *ACM Transactions on Computational Linguistics*, 2022.

[5] Google DeepMind. Google veo, 2024. https://deepmind.google/technologies/veo/.

[6] Brent Daniel Mittelstadt. Principles alone cannot guarantee ethical ai. *Nature Machine Intelligence*, 1(11):501–507, 2019.

[7] OpenAI. Gpt-4 technical report. OpenAI Research, 2023. `https://openai.com/research/gpt-4`.

[8] Aditya Patel, Bowen Chen, and OpenAI. Introducing dall-e 3. *OpenAI Blog*, 2023. https://openai.com/dall-e-3.

[9] Runway AI. Runwayml, 2024. https://runwayml.com.

[10] Ray Smith. An overview of the tesseract ocr engine. *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, 2007.

[11] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, et al. Transformers: State-of-the-art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2020.