

---

# Documentation of AI Street Interview

Aman Keskar

August 31, 2025

## 1 Introduction

This report documents the detailed and iterative process of creating a short, AI-generated video, specifically an "AI street interview," using a large language model (LLM) with video synthesis capabilities. The primary objective of this research task is to meticulously document the workflow, including the prompt engineering, qualitative analysis, and successive attempts at producing a desired creative artifact. This project builds directly upon previous research, such as the descriptive statistics and sports analysis work documented at [https://github.com/amankeskar/Task\\_05\\_Descriptive\\_Stats](https://github.com/amankeskar/Task_05_Descriptive_Stats), and extends that focus on data analysis into the domain of creative media production. The core hypothesis is that successful and nuanced AI content generation is not a single-step process, but rather a human-in-the-loop system that requires precise prompting and iterative refinement.

## 2 Methodology

The methodology for this project was centered on a process of iterative prompt engineering and qualitative analysis. The LLM served as the primary tool for content generation, with the researcher providing the creative direction and subsequent feedback. The workflow was executed in three distinct phases:

1. **Initial Prompt Formulation and Generation:** A comprehensive, multi-part prompt was crafted to establish the scene's visual and thematic elements. This included a detailed scene description, character profiles, a specific comedic tone, and a complete script. The generated video from this initial prompt served as the baseline artifact.
2. **Qualitative Analysis and Iterative Refinement:** The initial video was analyzed against the prompt's requirements. A critical focus was placed on the synthesis of "deep fake" elements, particularly the voice quality, which was a key aspect of the project's success. This analysis identified specific shortcomings that were then used to refine the next prompt.
3. **Final Generation and Conclusion:** The refined prompt was submitted to the LLM to produce an improved video. The success of this final attempt was used to draw conclusions about the most effective methods for AI-driven creative production and the importance of a detailed, iterative workflow.

---

### 3 Phase 1: Initial Prompt and Generation

The initial objective was to generate a short, humorous video with a clear comedic contrast. The creative concept was built on the juxtaposition of a fantastical, cheerful character and a realistic, serious one. The scene was set on top of a giant pyramid, with two distinct characters: a cheerful Pikachu interviewer wearing a tiny detective outfit and a serious, analytical gorilla in a professor's blazer. The tone was intended to be a lighthearted juxtaposition of a silly character asking a question and a serious character providing a factual answer. The following comprehensive prompt was submitted to the LLM to provide it with a clear creative brief:

A short, funny 8-second interview video. Scene: On top of a giant pyramid, with desert and sky in the background. Two characters sit with microphones. Character 1: Pikachu interviewer wearing a tiny Detective outfit, cheerful, says "Pika" sometimes. Character 2: A gorilla in a professor's blazer and glasses, serious and analytical. Tone: Comedic contrast. Pikachu asks silly questions, Gorilla gives a real stat. Script: [Pikachu]: "Pika pika? Who's the fastest in baseball?" [Gorilla]: "The fastest MLB players sprint just over 30 miles per hour." [Pikachu]: "Pika! Faster than me?" [Gorilla]: "Definitely. Numbers don't lie."

The LLM generated a video that successfully captured the visual elements and scripted dialogue. The characters and scene were rendered accurately, and the comedic timing of the script was preserved. However, a critical aspect of the "deep fake" or "AI street interview" was missing: the authenticity of the character voices. The Pikachu voice lacked its characteristic high-pitched squeak, sounding instead like a generic computer-generated voice. Similarly, the gorilla's voice was not sufficiently deep or resonant to enhance the comedic contrast, failing to convey the intended gravitas. This first attempt, while a functional proof of concept, served as a crucial baseline for the subsequent refinement process.

### 4 Phase 2: Iteration and Refinement of the Prompt

Based on the qualitative analysis of the initial generation, the next step in the workflow was to refine the prompt to specifically address the audio shortcomings. This demonstrates a key aspect of AI-driven creative work: the need for precise, descriptive prompts to guide the model toward a desired outcome. The updated prompt focused on specifying the vocal characteristics of each character to achieve the intended comedic effect.

The refined prompt included the following modifications:

- **Pikachu's voice:** Specified as "squeaky like the real Pikachu voice." This was a direct instruction to replicate a familiar and specific sound that is a core part of the character's identity.
- **Gorilla's voice:** Specified as "with the voice of Shaq." The use of a recognizable celebrity voice provided a clear and powerful aural reference, intended to heighten the comedic contrast between the serious subject matter and the surreal characters. This instruction moved the task from generic voice synthesis to a more complex "aural deep fake" challenge.

---

This deliberate act of prompt engineering was designed to leverage the LLM’s advanced synthesis capabilities, moving beyond simple dialogue generation to creating an audio track that grounded the surreal characters in familiar, archetypal sounds, thereby enhancing the overall comedic effect.

## 5 Phase 3: Final Generation and Analysis

The second attempt, using the refined prompt, yielded a significantly improved result. The new video successfully integrated the requested voice characteristics, creating a more polished and humorous final product. The contrast between the squeaky, high-pitched Pikachu and the deep, commanding “Shaq” voice of the gorilla created a more impactful comedic dynamic. This result validates the hypothesis that iterative prompting is essential for achieving nuanced, high-quality output from an LLM.

This research task demonstrates that creating a high-quality “deep fake” or “AI street interview” is a multi-step process that relies heavily on iterative feedback loops. The initial generation is often a starting point, and the quality of the final product is directly proportional to the detail and specificity of the subsequent prompts. The most valuable output of this process is not the final video itself, but the documentation of the workflow that led to its creation. This workflow provides a repeatable model for leveraging AI as a creative tool in future research endeavors.

## 6 Conclusion and Future Work

The project successfully documented a detailed workflow for using an LLM to generate a creative video artifact. The key finding is that successful AI content generation is not a single-step process, but an iterative one that requires the researcher to act as a prompt engineer, guiding the AI through successive refinements. The ability to provide specific, qualitative feedback, such as requesting a “squeaky” voice or a celebrity impersonation, is crucial for moving from a functional prototype to a polished, creatively aligned product.

For future work, this research can be expanded to explore more complex creative challenges. This could include:

- Experimenting with multi-speaker voice synthesis for more complex dialogues.
- Generating videos with more dynamic camera movements and scene changes.
- Exploring the use of AI for generating long-form narrative content.

The continued exploration of these workflows will be crucial for understanding the full potential and limitations of AI as a tool for media production.