INTERNSHIP PROJECT REPORT

Company: Elevate Labs

Project Title: AI Dungeon Pro: Multimodal GenAI Storyteller

Submitted By: Md Salman Farsi

---

1. ABSTRACT

The "AI Dungeon Pro" project is an interactive, artificial intelligence-powered text adventure game providing a multimodal storytelling experience. By leveraging Generative AI models, the application responds to user actions with dynamically generated text, contextual imagery, and voice narration. The objective was to integrate NLP, Computer Vision APIs, and Text-to-Speech into a single web application, demonstrating practical capabilities in creative writing.

2. INTRODUCTION

Generative AI has revolutionized user interaction. Traditional adventure games rely on pre-programmed decision trees, limiting freedom. This project utilizes a pre-trained Large Language Model (LLM) to generate story continuations based on open-ended inputs. To enhance immersion, it incorporates Stable Diffusion for real-time visualization and Google Text-to-Speech (gTTS) for audio. The application allows users to customize their experience across genres (Fantasy, Sci-Fi, Horror), serving as a prototype for next-generation media.

3. TOOLS USED

- **Programming Language:** Python
- **Frontend Framework:** Streamlit (interactive web UI)
- **Text Generation:** Hugging Face transformers (GPT-2 model)
- **Image Generation:** Hugging Face Inference API (Stable Diffusion v1.5)
- **Audio Synthesis & Utils:** gTTS (Text-to-Speech), requests, Pillow, re

4. STEPS INVOLVED IN BUILDING THE PROJECT

- **Environment Setup:** Developed a responsive, dark-themed web interface using Streamlit with custom CSS to align elements optimally. Added a sidebar for genre selection and parameter tuning.
- **Text Generation:** Loaded the pre-trained GPT-2 model. Designed a prompt-engineering template that dynamically incorporates the chosen genre, character, and action to generate a coherent story continuation, cleaned using Regex.
- **Image Generation:** Connected to Stable Diffusion v1.5 via API. Created a dynamic styling map that alters the image prompt based on the selected genre (e.g., adding "neon" for Sci-Fi), matching the narrative tone.
- **Voice Narration & Deployment:** Implemented gTTS to convert AI text into audio rendered on the frontend. Utilized Streamlit's session_state for chat history and added an export feature to download the log.

5. CONCLUSION

The project successfully demonstrates the seamless integration of text, image, and audio modalities into an engaging interface. By combining GPT-2, Stable Diffusion, and gTTS, the application showcases the potential of Generative AI in creating personalized experiences. Future iterations could involve fine-tuning larger models for longer context retention or local deployment of image models to reduce latency.