AI-Powered Immersive Kids Story Generation Task

Project Goal

Design, implement, and deploy a minimum viable product (MVP) that generates a complete, immersive children's story, including the final audio output combining a natural-sounding narration and complementary background sound effects/music.

Phase 1: Research and Documentation (Pre-Implementation)

The candidate must conduct and document thorough research on the following topics before writing any code. The final output of this phase is a detailed, professional **Research Documentation Report**.

Research Topics

1. **AI Story Generation Models:**
    - Compare the capabilities of at least three leading models (e.g., Gemini, Claude, OpenAI's GPT-4) for generating **children's stories**. Focus on features like **narrative coherence, age-appropriate content, emotional range, and "immersive" prompt-following** (e.g., "a story about a brave squirrel that feels scared when the wind blows").
    - Document the **prompt engineering techniques** required to consistently produce high-quality, natural stories suitable for reading aloud.
2. **AI Speech/TTS Generation:**
    - Compare and select a suitable third-party commercial service (e.g., ElevenLabs, Azure, Google Cloud TTS) for generating **natural, expressive narration**. Justify the choice based on **naturalness, emotion control (SSML support), child-friendly voice options, and API stability**.
    - Identify and document at least **three viable open-source alternatives** for Text-to-Speech (TTS) (e.g., Coqui TTS, Mozilla TTS, MaryTTS).
    - Analyze the **accuracy/quality** trade-off between the selected commercial service and the open-source alternatives.
3. **AI Sound Generation (Sound Effects/Ambiance):**
    - Investigate services (e.g., OpenAI, third-party libraries) or techniques for generating **contextual sound effects and ambient music** that align with the generated story's plot points (e.g., a "whoosh" sound for a flying scene, gentle music for a quiet moment).
4. **Scalability and Cost Analysis:**
    - Perform a detailed **cost analysis** for the end-to-end process (Story , speech, Sound) for **100,000 story generations** using the selected commercial services.
    - Propose a **cost-efficiency strategy** for managing this process at large scale, including potential hybrid solutions (e.g., using open-source TTS for lower-priority stories).

Phase 2: Implementation and Deployment (MVP)

The candidate will implement the end-to-end pipeline based on the findings from Phase 1.

Technical Requirements

1. **Story Generation Module:** Develop a script that uses an AI service (Gemini, Claude, or OpenAI) to generate a unique, natural-language, immersive children's story based on a simple input theme (e.g., "A fox who learns to share").
2. **Speech Generation Module:** Use the selected third-party TTS API to convert the generated story text into a **natural-sounding speech audio file (MP3 format)**. The candidate should demonstrate the use of features like **Speech Synthesis Markup Language (SSML)** to control voice, pitch, and pacing for a more "natural" delivery.
3. **Sound Integration Module:** Integrate a method (whether a simple library or a sophisticated AI tool) to generate and overlay relevant background sounds or music onto the speech track.
4. **Final Output:** Produce a **single, final MP3 file** combining the natural speech narration and the complementary background sound/music. Story length should be between 3 to 5 minutes. Required at least 3 stories.
5. **Code Repository and Licensing:**
    - Publish all source code on a **public GitHub repository**.
    - The repository must include a **MIT License** file.
    - The repository must contain clear documentation (README.md) on setup, usage, and dependency management.

Phase 3: Final Report and Critical Analysis

The candidate must compile the results into a **Final Project Report** that is straight to the point and avoids all AI-generated words in the report itself.

Report Structure

1. **Project Summary:** Brief overview of the final MVP and the technology stack used.
2. **Research Documentation (From Phase 1):** The full, detailed analysis of models, services, and alternatives.
3. **Implementation Justification:**
    - **Selection Rationale:** Clearly state *why* the chosen LLM for story, TTS service, and Sound Generation method were used, explicitly referencing the **trade-offs** analyzed in the research phase (e.g., "We chose ElevenLabs over Coqui TTS because its naturalness is higher and its API latency is milliseconds faster, justifying the cost per 1,000 characters.").
    - **Open-Source Alternatives:** Reiterate the identified open-source alternatives and explain the feasibility and challenges of migrating to them for cost savings.

4. **Accuracy and Quality Assessment:**
    - Qualitatively assess the **naturalness and appropriateness** of the generated story for a child audience.
    - Provide a quantitative/qualitative assessment of the **speech generation accuracy** (e.g., how often it mispronounces words, ability to convey emotion).
5. **Scalability and Cost-Efficiency Deep Dive:** Summarize the large-scale cost model and the strategy for optimizing operational expenses.
6. **References:** List all research papers, existing references, and tutorials consulted for the project.