University of Washington
Michael G. Foster School of Business
MSIS 549: ML and AI for Business Applications

Project Report
AI Storyteller

Team 1:
Yue Guo
Weishun Soong
Elena Tsaregorodtseva
Charis Wen
Li Li
Sneha Sahu

April 2024

1. Project Info

Our application generates stories for children of various age groups and displays them in a book format, complete with generated text and images, and also narrates these stories. Additionally, it includes the option of voice cloning for a personalized storytelling experience.

2. Project Scope

Our primary focus has been on delivering a compelling product pitch that outlines the application's features, target market, and potential for scalability and monetization.

The pitch is centered on introducing AI Storyteller as an innovative solution tailored for busy parents and their kids. It emphasizes the practical applications of the tool in everyday life and its potential benefits for educational institutions as well as for mental health support. The pitch strategy includes:

- Demonstrating the application's unique selling propositions, such as voice cloning and story generation along with image generation.
- Outlining strategic partnerships with educational and therapeutic organizations.
- Detailing a monetization model and future development phases to showcase growth potential.
- Emphasizing ethical considerations, prioritizing child safety and privacy.

3. Technical Aspects and Technology

The AI Storyteller application integrates several advanced technologies to deliver a seamless and interactive storytelling experience, enabling story generation, image creation, and voice cloning features.

AI Models and APIs:

- **OpenAI's GPT-4:** The core story generation component leverages OpenAI's GPT-4 model, utilizing its language understanding and text generation capabilities to create engaging and age-appropriate stories.
- **DALL-E 3:** For generating images that complement the story, we utilize OpenAI's DALL-E 3, a state-of-the-art image generation model. This model synthesizes images from textual descriptions, providing visual support to the narratives crafted by GPT-4.
- **OpenAI API:** This API facilitates interaction with OpenAI's suite of tools, including GPT-4 for text generation and DALL-E 3 for image creation.

- **ElevenLabs' Voice Cloning:** The voice cloning feature is powered by ElevenLabs' API, which allows for the realistic simulation of a user's voice. This technology is critical for providing a personalized experience, making it possible for parents to create voice clones that can narrate stories in their absence.
- **ElevenLabs API:** Integrated into our platform, this API is used for the voice cloning functionality, allowing us to convert text into audio with a specific voice signature.

Tools and Platforms:

- **Flask:** The application's backend is built using Flask, a micro web framework written in Python. Flask provides the necessary flexibility and scalability to handle web requests and serve the AI-generated content.
- **HTML/CSS:** The user interface is created using HTML5 and CSS, providing a responsive and accessible web experience. CSS frameworks like Bootstrap are utilized for rapid UI development, ensuring a consistent and mobile-friendly design.
- **JavaScript:** For dynamic content on the client side, JavaScript is used to enhance user interactions and provide a seamless experience within the browser.

4. Lessons Learned

What went well in our project was the robustness of the story generation based on user prompts, the high quality and contextual accuracy of the image generation, and the functionality allowing users to save their favorite stories or delete ones they did not like. These features combined to provide a highly customizable and user-friendly experience.

What could have been better in our project was reducing the time it took to generate stories. Faster story generation would have significantly enhanced user engagement by delivering instant gratification and a smoother, more responsive storytelling experience.

For challenges, one area we focused on improving was our prompt engineering for image generation. Initially, the images generated were sometimes irrelevant or inconsistent with the context of the stories. By refining the prompts and incorporating more specific descriptors and context, we were able to achieve a higher quality and relevance in the images produced, enhancing the overall coherence of the story visuals.

5. Future Development

The roadmap for AI Storyteller aims to enhance user engagement and broaden the application's capabilities. The following are key areas of focus for future development:

- **Adding an array of different voices for users to choose from.** We will integrate a selection of default voice options within the application. In addition to the pre-set options, users will have the ability to save and select cloned voices of family members, like parents or grandparents. This personalized touch allows for stories to be narrated in a familiar voice, maintaining the emotional connection, especially when a loved one is not present.
- **PDF download capability.** We will implement a feature that allows users to download stories in PDF format, enabling parents and children to enjoy AI-generated stories offline, providing a traditional reading experience with the convenience of portable, printable, and shareable content.
- **Inference speed optimization.** By optimizing our algorithms and enhancing our hardware capabilities, we aim to reduce wait times for story generation, image rendering, and voice synthesis, ensuring a near-instantaneous storytelling experience.
- **User authentication.** We will be implementing a robust user authentication system to ensure that user data is secure and private, featuring MFA processes to safeguard accounts and complying with global data protection regulations.
- **User-centric customizations and interactions.** Future iterations of the application will focus on deepening user interactions, including customizable character traits, plot directions, and the inclusion of personal elements within the stories, fostering a more interactive and immersive storytelling experience.

References

- Elevenlabs documentation
- Elevenlabs documentation on GitHub
- OpenAI documentation
- The app source code on GitHub
- Product features and system designs