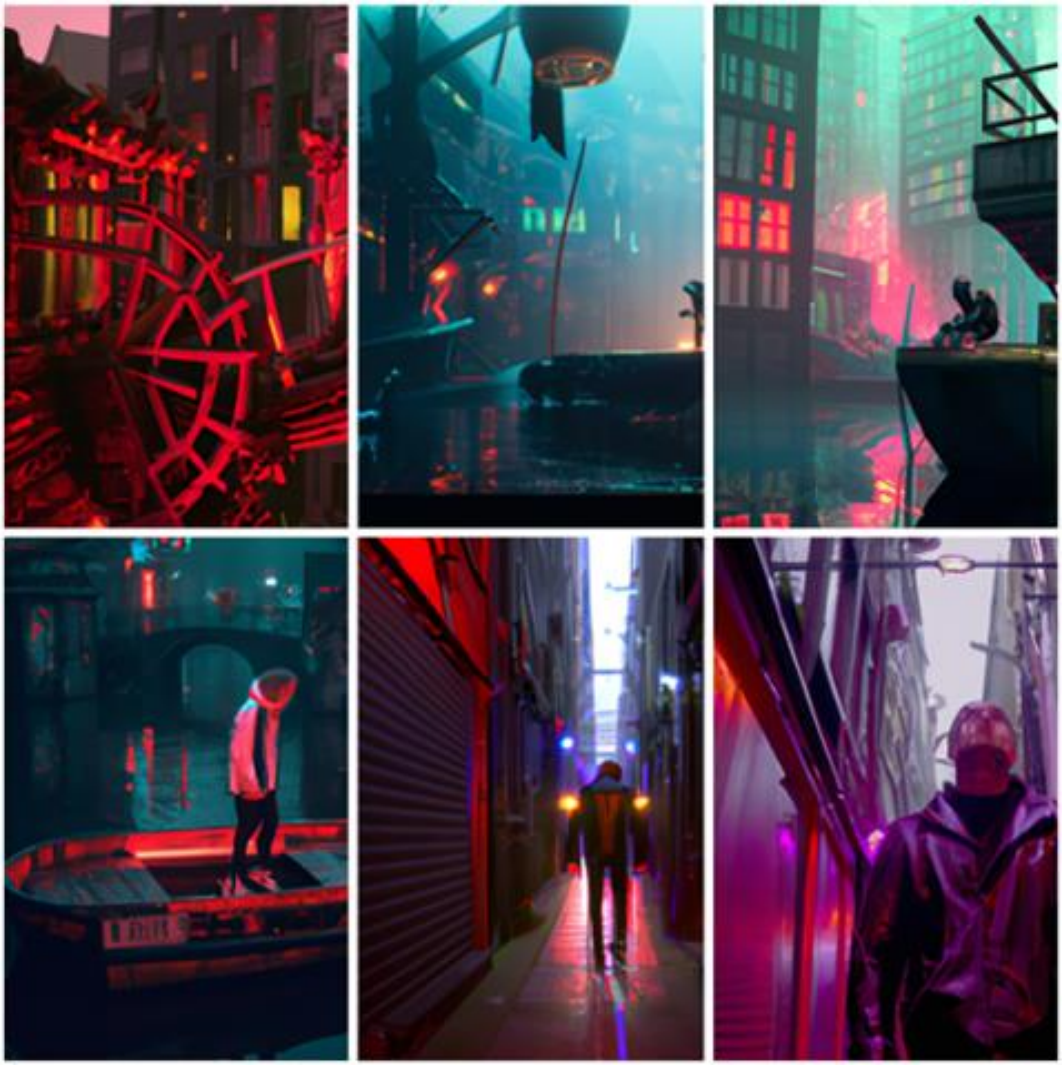
SAXION
UNIVERSITY OF
APPLIED SCIENCES



AI Storyboard Generator – Script Forge

CMGT: Graduation Thesis

Drilon Bajrami 465242

SAXION UNIVERSITY OF APPLIED SCIENCES

AI Storyboard Generator – Script Forge

Date 20 June 2023
Reference Graduation Thesis
Version 2.0
Department Creative Media & Game Technologies

Author Drilon Bajrami

Table of Contents

1 Abstract

This research paper provides a comprehensive exploration of the main topic for this graduation assignment, encompassing the clients' needs and requirements for a storyboard AI generator tool, research on various generative AI tools, the design process of the end-product, ideation and prototyping stages, and final testing and findings. The paper begins with an introduction to the main topic, establishing its relevance and significance. The clients' needs and requirements are then analyzed, identifying the key aspects that shape the project's direction. Extensive research is conducted on different generative AI tools, examining their capabilities, limitations, and suitability for the project's objectives. The design process of the end-product is detailed, outlining the steps taken to transform the clients' requirements into a tangible solution. This includes ideation and prototyping, where multiple concepts are generated and refined to achieve an optimal design. Rigorous testing procedures are employed to evaluate the functionality, usability, and performance of the final product. The paper concludes with the findings derived from the testing phase, highlighting key insights and observations. The research presented in this paper provides valuable insights into the development of generative AI tools and their application in addressing specific client needs, offering practical implications and potential future research directions in this domain.

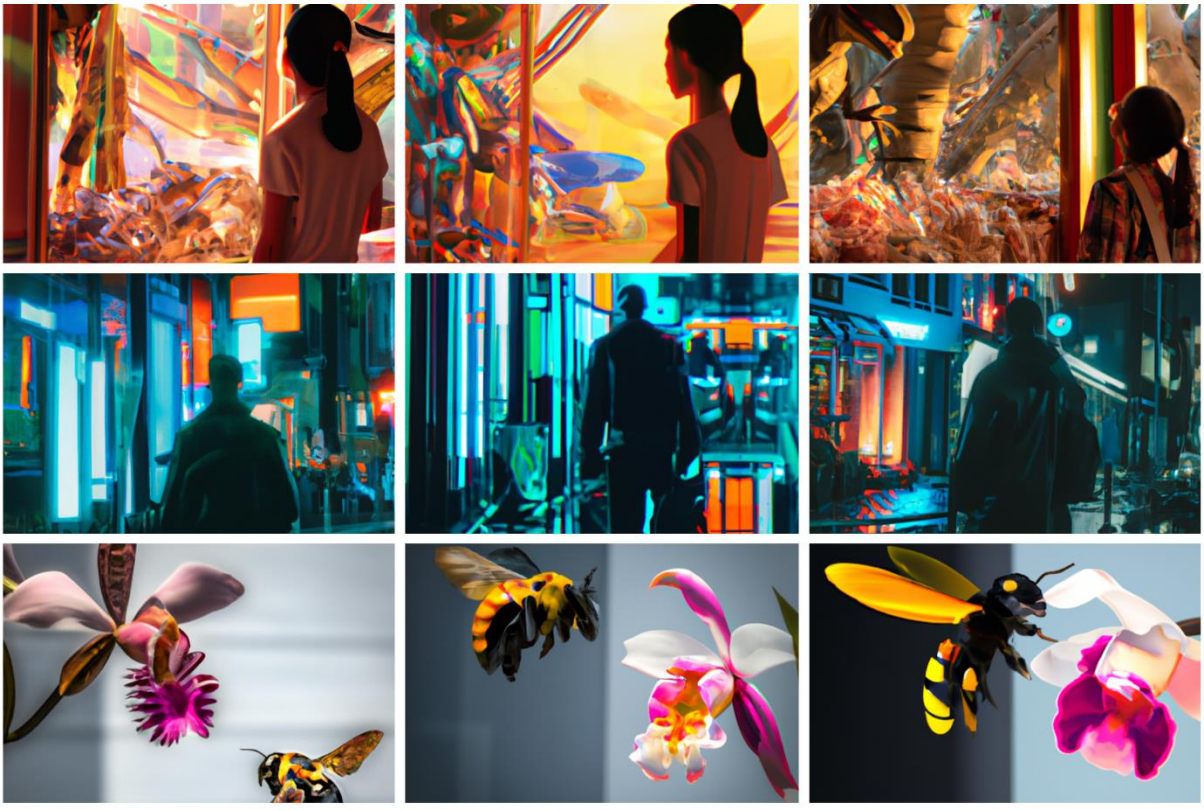


Figure 1 DALL-E 2 Generated Images and Variations

2 Introduction

2.1 Overview

For this graduation thesis internship, the given task is to develop a solution or product that can be used to create storyboards with images/videos made by existing generative AI tools (ex. Dall-E, Stable Diffusion, MidJourney, RunwayML, etc.). This solution is intended for the client, a non-profit organization, NPO (Nederlandse Publieke Omroep / Dutch Public Broadcasting). The end-product is expected to be used by content creators, who will generate visuals based on their content they want to pitch, in turn this will streamline the pitching process for the NPO workers, whom have the responsibility of evaluating pitched ideas and giving the green light on whether an idea should be processed further for development.

2.2 Graduation Project Team

This graduation project involves a team consisting of two students, namely the engineer (Drilon Bajrami - author of this report) and the designer (Christian Richters).
The engineer is responsible for various aspects of the development process, including selecting the technology stack for the end-product (prototype), implementing different functionalities, programming, conducting technical research on AI model's API libraries, performing user testing, and handling other technical aspects of the project.
On the other hand, the designer focuses on in-depth research of multiple existing generative AI tools, conceptualizing the desired end-product, conducting user tests, managing client communication, arranging meetings, and related responsibilities.
As a result, tasks are clearly divided between the team members, allowing everyone to apply their knowledge and experience in devising a solution that meets the client's requirements. This approach initiates a development process for the tooling, which can be further enhanced based on the final outcomes of this project.

3 Empathize

3.1 Company - NPO

The Dutch Public Broadcasting (NPO) is an independent administrative body comprising of multiple independent broadcasters that hold a broadcasting license.
NPO broadcasters are classified into two groups:

- “Broadcasters with legal functions in news, sports, education, and culture, such as NOS and NTR.
- Broadcasting associations that represent specific social or religious movements or target groups like young people and the elderly.” (About NPO, n.d.)

The broadcasters are accountable for the programmes they create, and their roles and obligations are legally specified, as are those of the NPO, the latter is responsible for coordinating the scheduling of all channels and online platforms and supervising the distribution and broadcasting of these programs. Additionally, NPO provides several general activities such as subtitling, rights management, procuring international programs, audience research, and promoting and marketing for all its channels and platforms. NPO is publicly funded and not authorized to earn profits from its services.
The NPO's mission is to "Open, to bring 18 million Dutch people closer together" (Our Mission, n.d.) while its values and objectives are to connect the Dutch public with programs that educate, inspire, and entertain with a diverse programming offer, ease of access, and transparency.
The NPO is owned by the Ministry of Education, Culture and Science (Dutch: Ministerie van Onderwijs, Cultuur en Wetenschappen; OCW), it was founded in the year 2000 and its headquarters are situated in Hilversum, Netherlands, with an approximate workforce of 450 individuals (Wikipedia-bijdragers, 2023). The official website for the NPO is accessible via this link: https://www.npostart.nl/.

3.2 End-user

The project targets NPO workers responsible for evaluating and selecting content ideas. They struggle with text-heavy scripts lacking visuals, hindering their understanding of the ideas pitched by different content creators. This uncertainty affects decision-making and raises concerns about production costs and viewer engagement.
To address this, the workers need an efficient way to evaluate pitches with visual elements. Incorporating visuals would save time, enhance comprehension, and increase the likelihood of selecting successful ideas.
With the rise of generative AI tools, the client sees an opportunity to improve their workflow. They have witnessed AI-generated visuals and seek a tool that enables content creators to easily generate visuals for their pitches, simplifying the evaluation process for the NPO workers themselves.
While ethical concerns exist, using generative AI tools for internal purposes alleviates these issues as no profit is generated from the content.
By adopting such a tool, the NPO workers can overcome their challenges, facilitating easier and more efficient evaluation of pitched ideas.

3.3 Market analysis

3.3.1 Market competition

Despite AI tools being relatively new, there is an abundance of image and video AI generator tools available on the web. While many of these tools utilize models developed by OpenAI and Stable Diffusion, some may use proprietary models that are tailored to their specific needs.
In this field, the main competitors include DALLE-2 from OpenAI, Stable Diffusion, Midjourney, Deep Dream, Nightcafe AI, and others. These generators are trained on models that use a vast amount of

data from the web, which enables them to produce higher quality images and generate content with various styles. While most of these generators require users to sign up with an email address, they are not entirely free, and some may offer short free trials or tokens for generating images. Some generators may renew the number of tokens available to users once every month.
As for other image AI generator tools that are less popular, many of them use AI developed by OpenAI or Stable Diffusion, through calling their APIs. These tools have their own trained models that are tailored to specific needs. For instance, some of these tools can generate only logos for brands, while others can create only cartoonish or anime-style content. There are also tools that generate short videos based on image input or text prompts. However, these tend to be more expensive since they generate many images behind the scenes, with the same seed but with slightly altered variables, such as camera movement, camera zoom, frame rate, shot length, and so on.
Training a custom model is a conceivable option, yet it is impracticable considering the resources required, such as a substantial amount of data with diverse styles and characteristics, as well as the expenses involved in generating said images. Moreover, due to limitations in terms of team size and expertise in this field, such an objective is deemed unrealistic. Thus, it is a better option to utilize existing tools instead.

3.3.2 Technological advancements

In recent years, the use and discussion of AI (Artificial Intelligence) have significantly increased across various fields. While AI tools have been in development for some time, they have only recently become widely accessible to the general public. Previously, these tools were limited to those in the technology, hardware, and software development fields. However, today, a growing number of people outside the technology field are utilizing AI tools for their own purposes and profit. Notable AI tools include large language models like ChatGPT-3 and DALL-E 2 from OpenAI, as well as MidJourney and Stable Diffusion.
The impact of these AI models is both positive and negative. On one hand, AI tools can enhance performance and streamline work processes for developers and content creators. However, ethical concerns arise regarding issues like image and code scraping from the network and generating content like copyrighted material. Additionally, there are concerns about job loss due to the automation of tasks previously performed by humans.
In the software engineering field, AI will not completely replace the workforce. Instead, it will create new opportunities and jobs, while improving productivity and reducing human errors caused by fatigue and stress. Although AI tools are highly "*intelligent*", they are not yet capable of designing and creating systems or solving complex problems that require human intervention. Therefore, AI should be regarded as a tool - a new skill to be acquired alongside others.

3.3.3 Case studies

The adoption of artificial intelligence (AI) has witnessed rapid growth across diverse industries, such as software development, data analysis, healthcare, retail, e-commerce, food, real estate, logistics, transportation, manufacturing, entertainment, music, gaming, and arts. AI tools have been tailored to address the specific requirements of each industry, offering solutions that enhance productivity, efficiency, and user experience.

One notable case study that incorporated text-to-image generative AI models is Canva, a web-based graphic design platform that allows individuals and companies to create designs quickly. While Canva incorporates an AI image generator tool based on Stable Diffusion, it is not specifically designed for generating storyboards for movies or documentaries. Instead, Canva focuses on facilitating drawing for individuals without design expertise. It offers features such as pre-designed templates, a drag-and-drop editor, an image and asset library, and brand consistency tools.

The integration of text-to-image generative models into the final product is a significant aspect. Given that pitches often involve substantial textual content, content creators can generate images using their own script as prompts, eliminating the need for exceptional drawing skills when creating storyboards to explain ideas to NPO workers. Furthermore, improving the user interface and consolidating various AI tools into a single web application would enhance their usability, as many of these tools lack clear user interfaces and may require time to comprehend their functionality.

4 Definition

4.1 Problem Definition

The present situation involves NPO workers as the target audience, along with content creators who pitch content ideas to them. The NPO workers are burdened with the task of evaluating numerous content ideas put forth by various content creators, leading to substantial time spent on reading extensive text scripts. Consequently, they face challenges in fully comprehending the essence and sentiment of the pitched content ideas. Moreover, the current pitching process is prolonged and incurs expenses. Compounding the issue, not all content creators possess the ability to create storyboards due to limited drawing skills or lack of experience with software used for creating visuals. This scenario highlights the need to address these challenges and streamline the pitching process to enhance efficiency and cost-effectiveness while ensuring clear communication of content ideas between the creators and the NPO workers.

4.2 Main Question

How to streamline the content pitching process, through a web application that bundles multiple different generative AI tools and keeps consistent styling, which can be used by the content creators to generate visuals for their content and consequently easing the pitch reviewing process for the NPO workers.

4.3 Sub Questions

1. What programming language and frameworks should be used to set up the back-end of the web application?
2. What JavaScript framework should be used for building the front-end application user interface?
3. How to set up a connection and communicate between the back and front-end sides of the web application?
4. Which AI tools offer the best documentation and easy to use API library, that can be implement in the back-end side of the web application?
5. What can be improved based on the usability of the existing generative AI tools?

4.4 Methodology

Sub questions:

1. **What programming language and frameworks should be used to set up the back-end of the web application and why?**
   For determining the suitable programming language and frameworks for the web application back-end, valuable insights from diverse sources such as online forums, tutorial websites, and developer communities are gathered. This qualitative data, derived from reviews, opinions, and recommendations shared by developers, helps in selecting the most suitable options based on trends and common recommendations.
2. **What JavaScript framework should be used for building the front-end application user interface?**
   To identify the recommended JavaScript framework for building the front-end user interface, valuable insights from online forums, tutorial websites, developer communities, and official documentation are collected. This qualitative data, obtained from reviews, opinions, and recommendations shared by developers, assists in selecting the most suitable framework by analyzing qualitative information and considering specific requirements.
3. **How to set up a connection and communicate between the back and front-end sides of the web application?**

Valuable insights on setting up a connection and communication between the back-end and front-end of the web application are obtained from various sources such as online documentation, tutorials, developer forums, and community resources. This qualitative data, gathered through active engagement with online communities, reviewing tutorials, exploring forums, and studying documentation, contributes to understanding the steps and best practices involved in establishing the necessary connection and communication.

4. **Which AI tools offer the best documentation and easy to use API library, that can be implement in the back-end side of the web application?**
   To determine which AI tools, offer the best documentation and easy-to-use API library for implementation on the back-end side of the web application, a combination of existing data sources (AI tool documentation, developer forums, technical articles, and online resources) and primary data collection (testing and evaluating documentation and API libraries) is utilized. This qualitative data helps in identifying AI tools with high-quality documentation and user-friendly API libraries, ensuring a smooth and efficient development process while enhancing the overall user experience of the web application.
5. **What can be improved based on the usability of the existing generative AI tools?**
   For improving the user interface of existing generative AI tools and implementing them in the web app, a combination of existing data sources (user feedback, UI/UX design principles, case studies, and technical articles) and primary data collection (user surveys, usability testing, and analysing user interactions) is used. This mix of qualitative and quantitative data provides insights into user preferences, pain points, and opportunities for improvement. The results inform decision-making to enhance the user interface, ensuring a more intuitive, user-friendly, and visually appealing experience aligned with the needs and expectations of web app users.

4.5 Scope

Deliverables: A web application that incorporate three distinct AI tools capable of converting text into images or videos for storyboarding purposes. Additionally, the application will have a user-friendly interface designed to ensure easy utilization by the end user.
Inclusions/Exclusions: Within the framework of this project, the team will conduct an examination of existing generative AI tools and explore the relevant research pertaining to the chosen technology stack for the web application's development. It is important to note, however, that the final application will not incorporate advanced image/video editing features. This decision is primarily based on time limitations and the limited size of the project team.
Assumptions: It is assumed that AI will be included in every possible field to improve the quality of work, along with making many jobs position go useless however introducing new and more specialized job positions. Moreover, there will be a lot of ethical issues with using AI, due to how the AI is trained.
Constraints: Due to the limited team size consisting of only one engineer and one designer, as well as a short development period for the end-product, the research project will face constraints. As a result, the focus will primarily be on researching the crucial aspects of AI and its implementation in an application for generating storyboards. Additionally, the engineer's lack of experience in the web technology stack will limit the features of the end-product, particularly in terms of extensive image/video editing tools.
Resource Allocation: The research project necessitates a substantial amount of user testing to identify any issues or deficiencies in the developed prototype. Additionally, it aims to uncover unforeseen issues arising from the unique nature of AI generative tools.
Budget: The project budget will be capped at 3,500 euros, covering the personnel expenses for a six-month duration. Additionally, approximately 400 euros from the total budget will be allocated for purchasing tokens required for the AI tools utilized. On average, the monthly expenditure for utilizing these AI tools will amount to 50 euros.

4.6 Research Goals

1. To create a web application that implements at least one text-to-image model, one text-to-video model and one model for character creation.
2. To learn how to develop both the back-end and front-end of a web-based application.
3. To improve and ease the use of the implemented AI tools, through a more comprehensible user interface.
4. To ensure that the generated content adheres to a specific styling scope while allowing users to generate content within those defined style boundaries.

4.7 Client requirements – Indicators of success

The client's questions regarding the storyboard generator solution, are:
- Can we keep consistent styles in a storyboard?
- Can we integrate famous people in a storyboard?
- Can we generate multiple variation for a single frame?
- Can we generate animations or videos?

5 Ideation

5.1 Development Environment Tools

5.1.1 Back-end server development tools and frameworks

After researching back-end server development tools and frameworks for web applications, promising options with ample documentation and popularity in developer communities and forums include Express.js, Django, Ruby on Rails, Flask, and ASP.NET. Here's a brief explanation of each framework and its pros and cons:

- **Express.js**: A fast and minimalist web application framework for building RESTful APIs with Node.js. It is widely used and considered the standard back-end server framework for Node.js. Pros include a simple and intuitive API and a vast ecosystem of middleware and plugins. However, being minimalist, it requires additional libraries for features like authentication and database integration, which can increase development time and complexity.
- **Django**: A high-level Python web framework following the model-view-controller (MVC) architectural pattern. It is known for its robustness, security features, and built-in functionalities, making it suitable for content-heavy and data-driven platforms. Pros include a full-featured and batteries-included framework with built-in ORM, authentication, and admin interface. However, it may be overkill for small projects and has a steeper learning curve for developers new to Python.
- **Ruby on Rails**: Also known as Rails, it is a web application framework written in Ruby. Rails' convention-over-configuration approach and Ruby's elegant syntax enable efficient development of feature-rich platforms. Pros include convention-based configuration, developer productivity, and built-in features like ORM, routing, and testing. However, Rails may have performance issues for high-traffic applications and requires familiarity with the Ruby language.
- **Flask**: A lightweight and flexible web framework for Python, designed to be minimalistic and provide essential tools for building web applications. Pros include ease of setup and getting started, making it suitable for small to medium-sized projects. It allows developers to choose additional libraries as needed. However, Flask's minimalistic approach may require more manual configuration compared to other frameworks and additional libraries for features like database integration and user authentication.
- **ASP.NET**: A web application framework developed by Microsoft for building enterprise-level applications using C# or VB.NET. Pros include a comprehensive framework with a rich set of tools and libraries, strong security features, and excellent scalability support. However, it can be more complex to set up and configure compared to other frameworks, primarily suited for Windows environments, and may not be the best choice for cross-platform or non-Microsoft projects.

Considering the choice of Python for the back-end of the web application, two viable options remain: Django and Flask. After comparing the use cases, needs, and requirements of the application, Flask emerges as the preferred option for developing the back-end server, which will integrate multiple AIs and their respective API calls. Flask allows for easily setting up a centralized management of AI platform accounts and the loading of API tokens.

5.1.2 Front-end user interface development tools

Front-end user interface development tools can greatly impact the success of a web application. Here are some popular options:

- **React**: A JavaScript library developed by Facebook, known for its component-based architecture, virtual DOM, and efficient rendering. React is used by Facebook, Instagram, WhatsApp, Netflix, and Airbnb. Pros include reusability, performance optimizations, and a vast

ecosystem. However, it has a steeper learning curve and requires additional libraries for routing and state management.

- **Angular**: A TypeScript-based framework developed by Google, offering comprehensive features like data binding, dependency injection, and routing. Angular is used by Google, Gmail, Microsoft Office 365, IBM Watson, and Autodesk Fusion 360. Pros include a structured approach and extensive tooling. However, it has a higher learning curve and increased complexity for smaller projects.
- **Vue.js**: A progressive JavaScript framework known for simplicity and ease of adoption. It allows for component-based UI development and is used by Alibaba, Xiaomi, Baidu, and Laravel. Vue.js offers intuitive syntax, reactivity, and excellent performance. Pros include a gentle learning curve and a supportive community. However, it may have a smaller ecosystem and lack some advanced features.
- **Bootstrap**: A widely used CSS framework providing pre-styled components and responsive design utilities. Twitter, Spotify, Airbnb, Udemy, and Slack utilize Bootstrap for front-end interfaces. Pros include quick creation of visually appealing and responsive designs. However, websites may have a similar look, customization requires effort, and file sizes can be larger.
- **Tailwind CSS**: A utility-first CSS framework offering flexibility and customization options. Tailwind CSS allows for rapid building of unique designs using utility classes. Pros include flexibility, customization, and excellent performance. However, it requires more setup and relies heavily on utility classes.

Considering popularity and usability, Vue.js emerges as a promising option for front-end development. It covers most web application requirements, is easier to learn, and offers flexibility for developers.

5.1.3 Communication between back and front ends

The communication between the front-end (*Vue.js*) and back-end (*Flask* + *Python*) components generally happens over HTTPs. These HTTP request will be made by using the *Axios* library.
An overview of the communication process is as follows:

1. Frontend: End users provide input prompts and prompt options in the Vue.js frontend. This input data is bundled and sent as an HTTP request to the backend.
2. Backend: The Flask + Python backend receives the HTTP request, extracts the bundled input data, and prepares a data object for making API calls to the selected AI tool.
3. API Calls: The backend uses the prepared data object to call the AI tools' API endpoints, requesting image or video generation based on the provided prompt data.
4. Response Processing: Upon receiving the response from the AI tools' API, the backend processes it to extract relevant data, such as the URLs of the generated images or videos. If the requests fail due to reasons, this will be relayed to the frontend as well.
5. HTTP Response: The backend generates an HTTP response containing the extracted data, specifically the URLs of the generated content.
6. Frontend Update: The Vue.js frontend receives the HTTP response from the backend and updates the user interface, accordingly, allowing users to access and view the generated images or videos.

By following these steps, users can input prompts and options in the frontend, which are then sent to the backend. The backend processes the data, makes API calls to AI tools, retrieves the generated content, and sends the URLs back to the frontend for display.

5.2 Choosing the right AI Tools

The research for selecting AI tools for the storyboard generator application was primarily led by designer Christian, with input from the engineer on technical considerations. Factors such as comprehensive API documentation and ease of implementation were considered.

It's important to note that limited research could be conducted due to the novelty of the field, resulting in fewer available tools for comparison. Interacting with these tools typically involves providing a prompt and generating output, though some tools may have more complex input parameters. Simplifying parameter selection is crucial for improving user-friendliness and enhancing the usability of the tools.

5.2.1 Text-To-Image

The realm of AI-based "text-to-image" models offers a diverse range of options, surpassing other models such as "text-to-video" in terms of availability and versatility. This study aims to conduct a fair comparison among the prominent text-to-image models, including DALL-E 2, Midjourney, and DreamStudio, to determine the most suitable model for integration into our application. Each model's capabilities were thoroughly researched, and this report presents the findings and conclusions drawn from the analysis.

Analysis

DALL-E 2 (*OpenAI*)
Compared to other models, DALL-E 2 stands out as a user-friendly and cost-effective choice. However, it may not produce photorealistic results as extensively as Midjourney and DreamStudio, although this can be achieved by utilizing specific prompting techniques that leverage different styles. The model's API documentation is comprehensive and user-friendly, making integration straightforward. Additionally, DALL-E 2 can be seamlessly integrated with ChatGPT using a single account or API token. Nevertheless, it is important to note that DALL-E 2 has certain limitations, such as restrictions on generating images of famous personalities, explicit content, violence, and political issues.

Midjourney (*Midjourney, Inc*.)
Midjourney stands out with its outstanding image quality and freedom from significant safety constraints. It enables the generation of images featuring famous individuals and diverse content. However, its usage is confined to the Discord platform, and the generated images may contain watermarks or signatures from the original artists, potentially raising copyright infringement concerns. Furthermore, all images generated through Midjourney are publicly accessible by default, providing limited privacy.

DreamStudio Stable Diffusion (*Stability AI*)
DreamStudio Stable Diffusion is recognized for its robust image generation capabilities, providing users with a wide range of customization options. While it remains cost-effective, mastering this model may require a significant learning curve due to the complex array of prompting choices available.

Conclusion
After careful evaluation, DALL-E 2 is determined to be the most suitable choice for integration into the web app. Its seamless implementation and comprehensive API documentation make it an ideal fit. This decision considers the strengths and limitations of DALL-E 2, alongside the complementary ChatGPT model, ensuring a well-informed decision-making process for the application.

5.2.2 Text-To-Video

Due to the limited availability of text-to-video models, in-depth research on their technical aspects was constrained. However, the decision to choose Deforum was made relatively quickly, which proved to be a suitable choice due to its extensive prompting options. Among the alternatives, Deforum (Stable Diffusion) emerged as the most fitting option, offering well-documented API calls and a wide range of features for prompt creation, such as camera movement, zooming, panning, and rotation. To enhance user interaction, an improvement could be made by incorporating intuitive sliders and user interface elements that streamline the input process for video generation, reducing reliance on textual descriptions and symbols.

5.2.3 Text-To-Character

When considering the text-to-character (speech) model, two primary options stand out: D-ID and synthesia.io. After thorough research and careful evaluation, D-ID was determined to be the optimal choice for this use case, primarily due to its utilization of Stable Diffusion technology. In comparison, D-ID offers a superior and more accessible API, simplifying its integration into the web app.
While synthesia.io focuses on generating realistic human videos, D-ID provides a broader range of prompting options. This versatility allows for the creation of customized voices, expanding the creative potential of the generated content. Additionally, the inclusion of Stable Diffusion technology in D-ID enhances the quality of the generated voices, resulting in a more immersive and satisfying user experience.

5.3 Concept Art & Storyboards

5.3.1 DALL-E 2 Preview

The existing DALL-E 2 official web application offers a user-friendly interface with a single input field for text prompts. Additionally, it provides an editing tool that allows users to modify images by filling in cut-out parts with generated content. However, due to the limited expertise in web development within the project's scope, the editing and filling feature will not be included in this implementation.

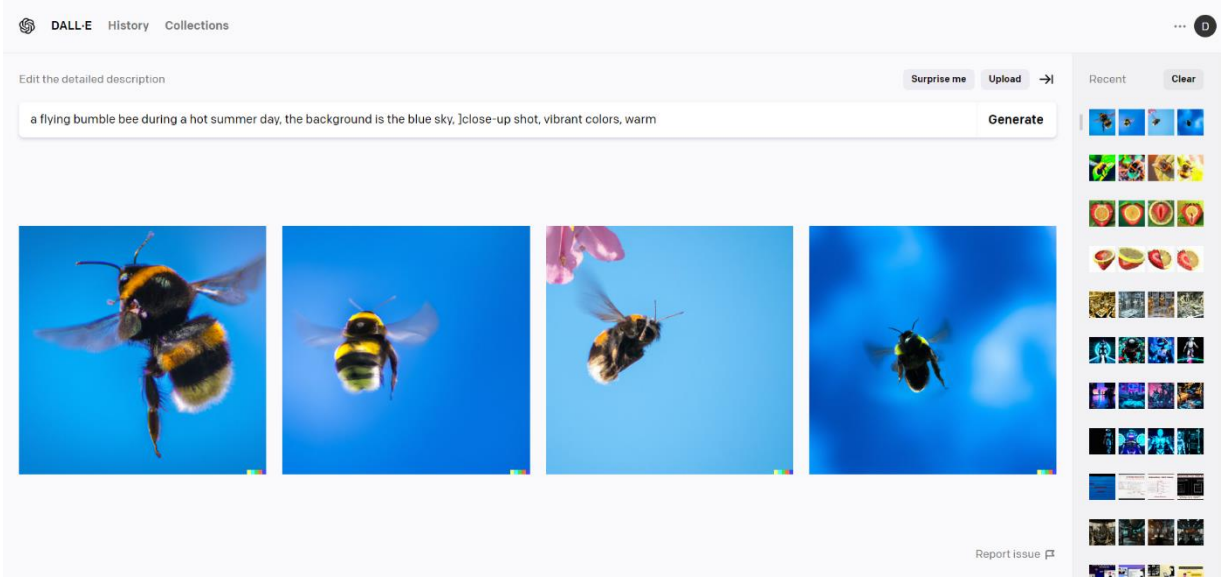


Figure 5.1 DALL-E 2 - Main Page

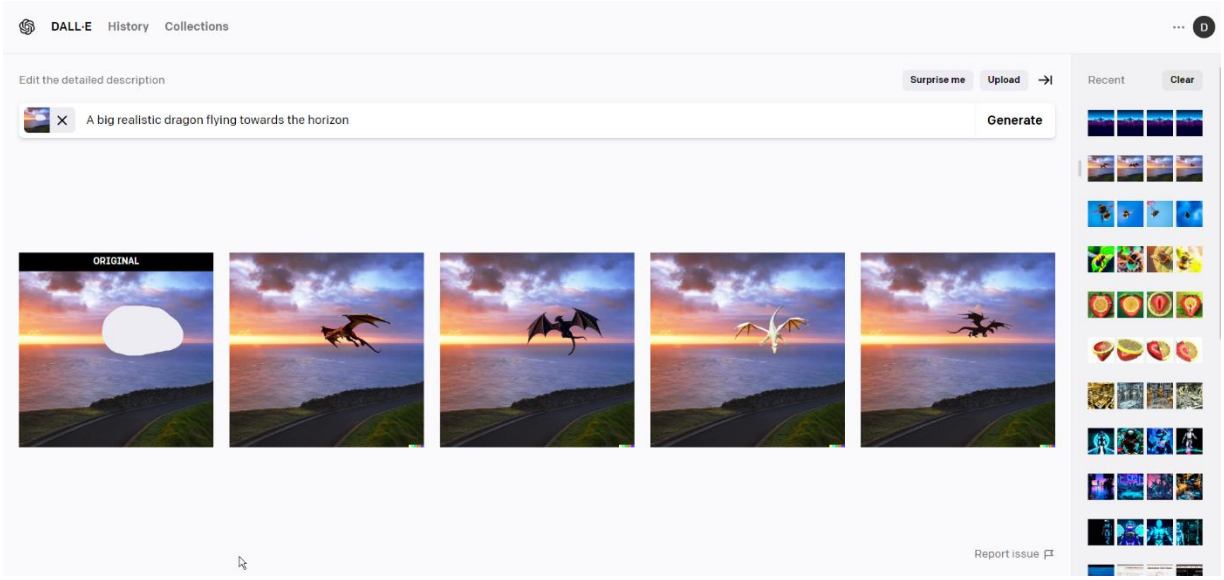


Figure 5.2 DALL-E 2 Image Edit Page

The final output of the DALL-E 2 model is influenced by the clarity of the prompt provided and the chosen prompt styling key words. It is important to note that even the creators of the DALL-E 2 model do not have a complete understanding of how it precisely interprets text prompts. However, the community has made significant contributions by sharing their generated content using their own designed prompts. An invaluable resource in this regard is the "DALL-E 2 prompt book," which has been created based on numerous prompts and content generated by the community. This prompt book covers various aspects of prompt styling, including emotions, artistic styles, camera angles, proximity and position, lenses, lighting, film types, TV prompts, illustration styles, places and spaces, materials, and more.

“DALL-E knows a lot about everything, so the deeper your knowledge of the requisite jargon, the more detailed the results.” (The DALL·E 2 Prompt Book, n.d.-a, p. 12)

5.3.2 Concept

To ensure the generated content aligns with the desired style, an additional section for prompt styling options is introduced alongside the initial prompt input field. Moreover, the implementation of the image variation feature aims to facilitate adherence to the chosen style preference of the end-user. These features are conveniently accessible on a single page, eliminating the need for navigating through separate pages to create image variations and set styling options.
Furthermore, ChatGPT has been integrated into the prototype to enhance the user's prompt input. This addition allows for improvements or rephrasing of prompts if the user lacks confidence in their personally written prompt and selected options.

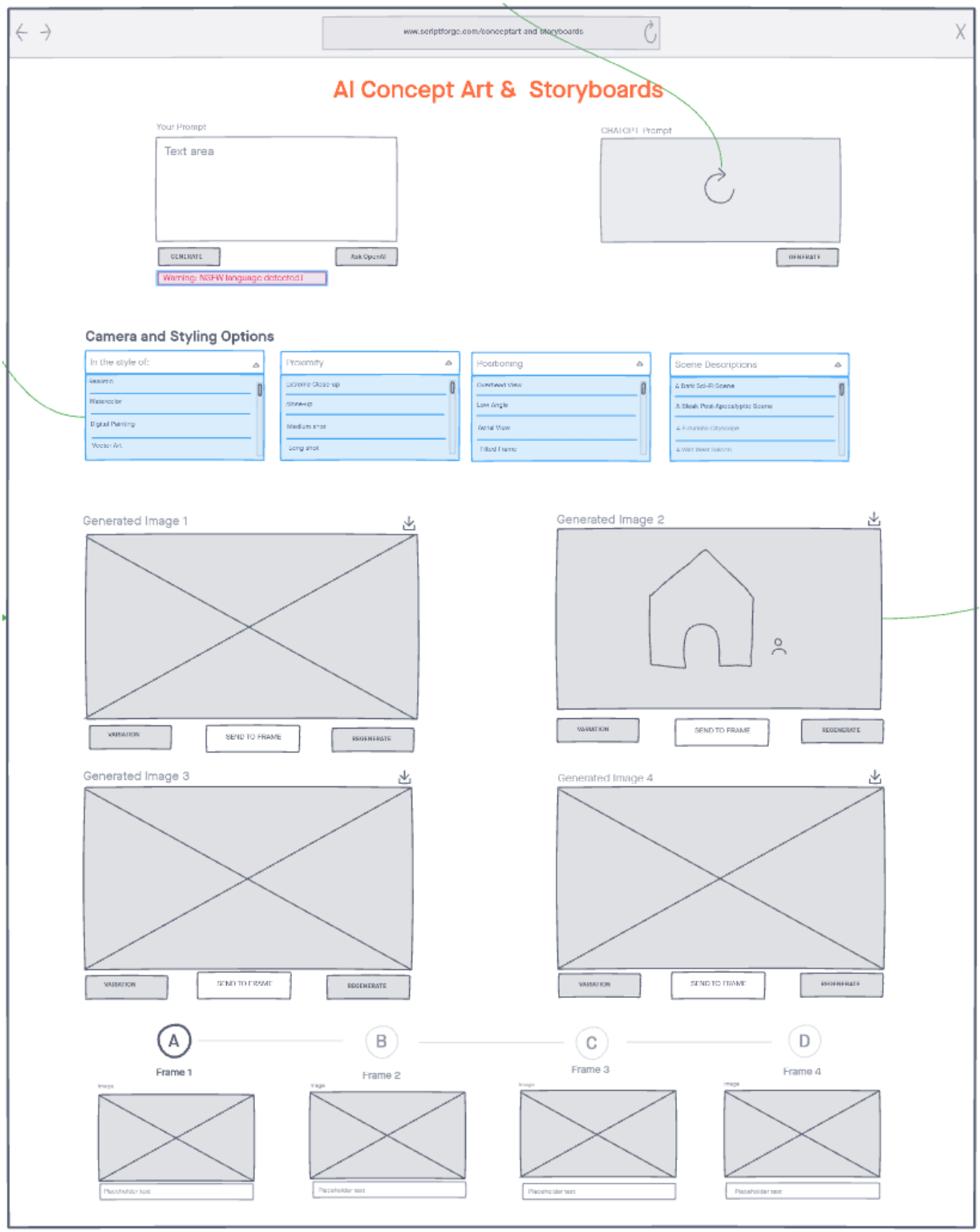


Figure 5.3 Concept Art & Storyboard Concept by Christian (Designer)

5.4 Videos & Animations

5.4.1 Deforum SD Preview

The current Deforum SD web application can be perceived as complex due to the presence of multiple animation prompt option fields, which can make prompt creation less user-friendly. Initially, there can be multiple prompts for the final video, each requiring specification of the starting frame number and camera movement options. To designate multiple options, users must separate each prompt using a pipe symbol (|), which introduces the possibility of errors if not done carefully.

For example, an animation prompt would follow this format: "0: Bag End Hobbiton | 100: Rivendell | 200: Tower of Cirith Ungol | 300: Mount Doom". Each animation prompt starts with the starting frame number, followed by a colon and a description. Multiple animation prompts are separated by the pipe symbol.

Additionally, the length of the generated video depends on the provided "fps" (frames per second) and "max_frames" parameters. Users must manually calculate the duration of the video and the duration of each animation prompt, which can lead to confusion. Similarly, camera movement details need to be written down with the starting frame number, desired value, and pipe symbol to separate different prompts.

It is important to note that Deforum offers camera movements in both 2D, and 3D space. However, for the sake of simplifying the user experience, the decision was made to focus solely on 2D camera movements.

Figure 4 Deforum SD Main Page

5.4.2 Concept

The team designer opted for a more user-friendly UI design, which enhances the clarity and separation of animation prompts. Additionally, camera movements are displayed as sliders with predefined ranges of possible values, reducing the likelihood of errors.

The "fps" (frames per second) has been set at a fixed value of 15. This decision was made based on the understanding that higher frame rates are not necessary for AI-generated animations. The focus is on conveying emotions or moods rather than relying on high-quality visuals or frame rates, which would lengthen the generation process and increase costs.

To provide a clearer understanding of the video duration, a slider will be implemented. This allows users to visualize the length of the generated video in seconds. The user interface will also include an element for selecting the duration in seconds or frame count for each prompt. This feature enhances

clarity regarding the duration of each animation prompt, and users can adjust the duration by sliding the dividers within the designated box.

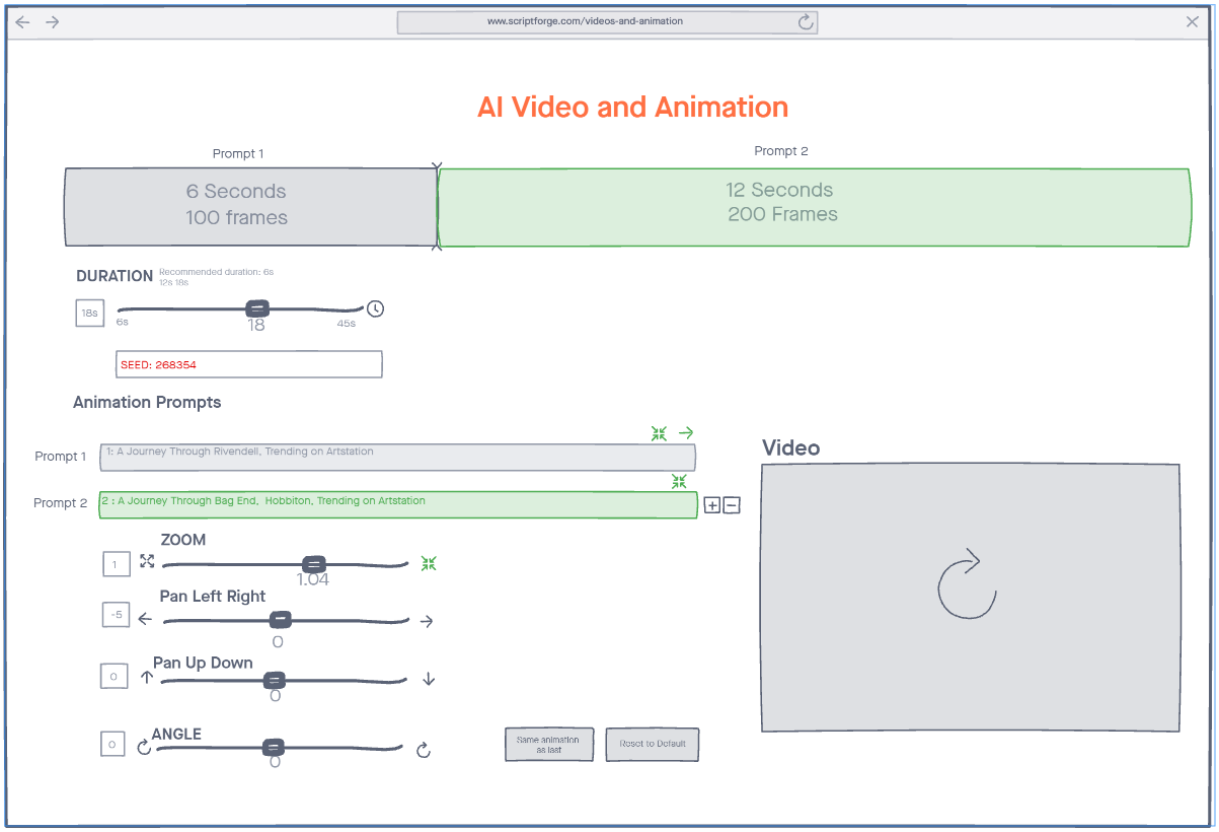


Figure 5.5 Video & Animation Concept by Christian (Designer)

5.5 Character Creation

5.5.1 D-ID Preview

The D-ID web app offers a user-friendly interface, providing straightforward usability. Users have the option to select a character from the existing ones or create their own. Additionally, the app features a text area for entering the dialogue prompt.

Moreover, the app allows users to customize various aspects of the dialogue. Users can choose the dialogue language, select a voice actor, and determine the voice style. Alternatively, users have the option to upload their own custom voice if desired.

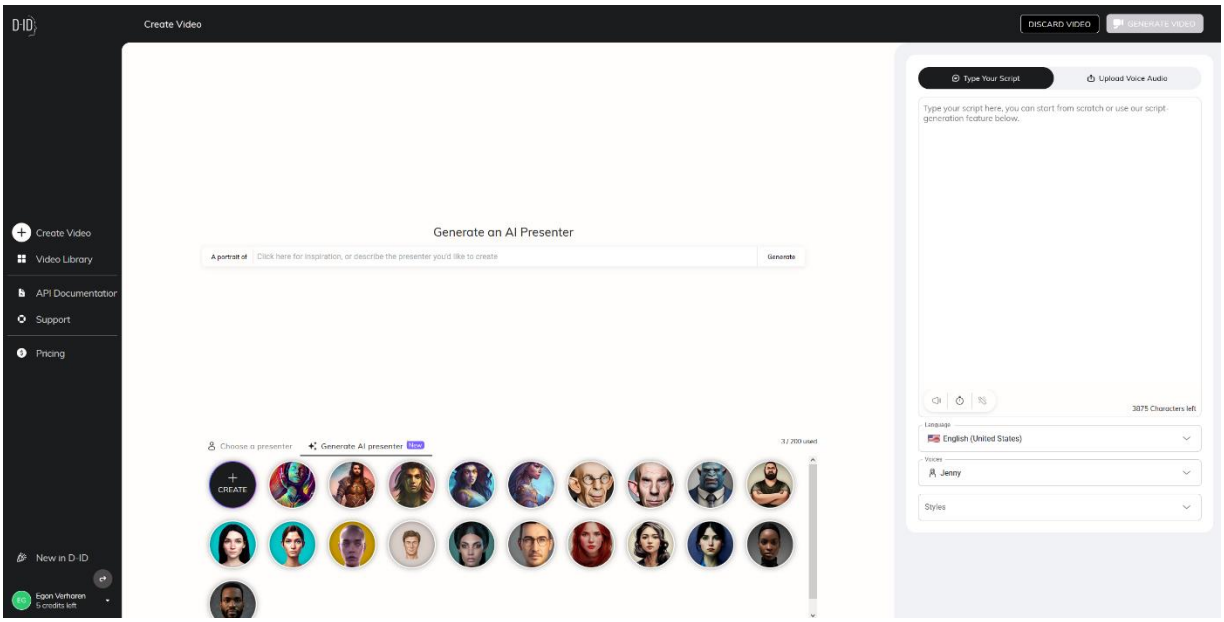


Figure 5.6 D-ID Main Page

5.5.2 Concept

In comparison to the text-to-image and text-to-video models, this model already possesses a user-friendly and intuitive UI. Consequently, the core concept of the original app will remain largely unchanged. The only notable addition is the integration of ChatGPT to enhance the prompts for character creation or for the dialogue.

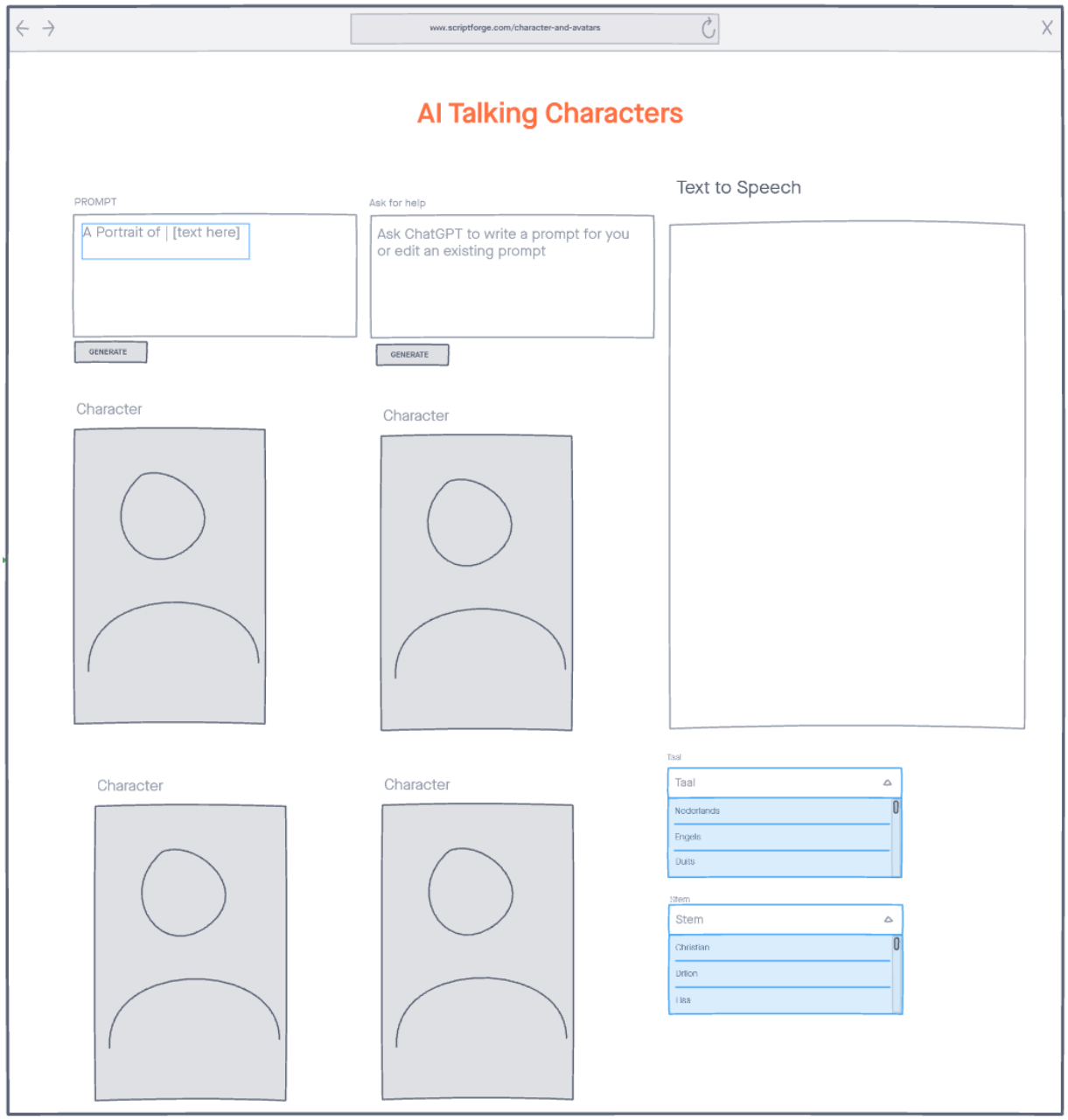


Figure 5.7 Character & Avatars Concept by Christian (Designer)

6 Prototyping

6.1 Concept Art & Storyboards

The focus in creating the prototype for the concept art page is to design a user interface that is both comprehensible and simple in terms of prompt input options.

6.1.1 First Prototype

The initial prototype was built with a focus on the back-end server using Python and Flask, with minimal integration of HTML and CSS for the layout. It followed a simple design approach, consisting of an input field and a "Generate" button. The main goal of this prototype was to establish the initial API call to DALL-E 2, retrieve and display the generated images on the page, and conduct preliminary testing with a limited group of users. This testing phase aimed to collect valuable feedback that would inform further iterations or the final version of the prototype.

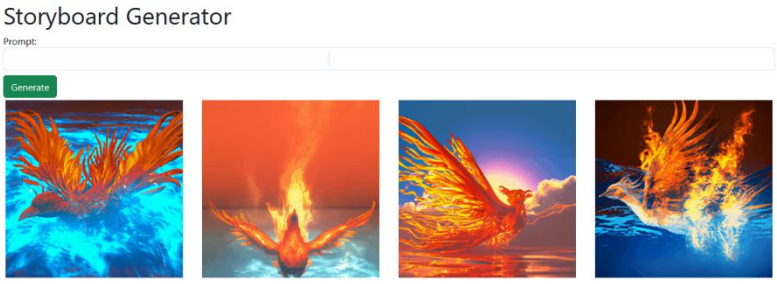


Figure 6.1 First Iteration of Concept Arts & Storyboards Prototype, DALL-E 2 implementation (Python + Flask)

User tests

In the initial prototype of the Concept Art & Storyboards tool, user testing was conducted to gather feedback and insights, despite its limited functionality. The user tests, which involved a small group of colleagues or peers, provided valuable input regarding the capabilities of image generator tools like DALL-E 2.

With the feedback received, the development process was initiated to incorporate additional functionalities that would allow users to generate content within specific style parameters. This iterative approach enabled the team to evaluate the feasibility and usefulness of integrating DALL-E 2 into the final solution.

6.1.2 Second Prototype

The second prototype of the Concept Art & Storyboards tool was developed exclusively on the front-end using the Vue.js framework library. It includes all the essential prompt input options such as the user input prompt text box, the ChatGPT text box for refined prompts, the styling options, and the display of generated images. Each generated image is accompanied by buttons for creating variations or initiating regenerations. However, it's important to note that in this iteration, the "Variation" and "Regenerate" buttons are not yet functional. The focus of this prototype is to test the effectiveness of

the styling options and the ChatGPT prompt refinement. The functionality of the "Variation" and "Regenerate" buttons will be implemented in subsequent iterations of the prototype.

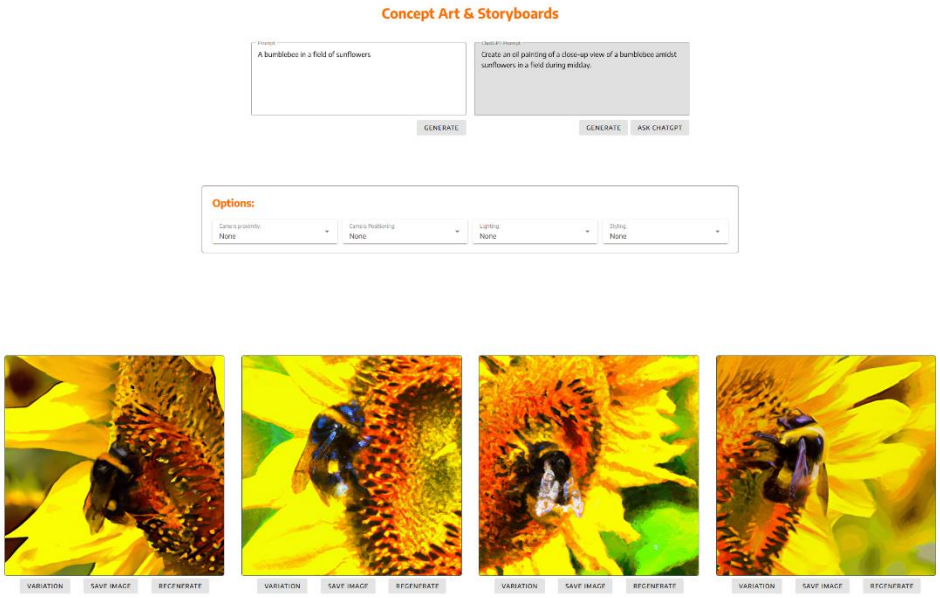


Figure 6.2 Second Iteration of Concept Art & Storyboards Prototype

User tests

The testing environment for the Concept Art & Storyboards tool included the web application itself, along with a Google questionnaire to collect user feedback. The testing sessions involved the presence of both the engineer and the designer from the team to provide assistance and gather insights. The engineer's role was to address technical difficulties, note any bugs or errors encountered, and use this information to make necessary improvements.
Eight users, including colleagues and peers associated with the graduation team, participated in the testing sessions. Their feedback and input were valuable in assessing the user experience, identifying areas for enhancement, and validating the functionality of the tool.

User Tests Feedback Summary

The user testing feedback collected from the sessions has been filtered by the engineer to concentrate specifically on the technical aspects of the web application prototype. This selective approach ensures that the feedback included in this report is focused on improving the next iteration of the prototype. Including all the results and feedback in the report could make it overly complex and potentially distract from the key technical insights. However, for a comprehensive record of the user feedback, complete test forms and response summaries, along with the source code, are available in the Graduation Dossier.

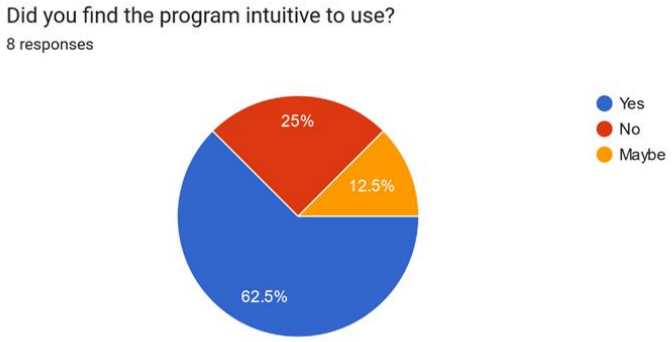


Figure 6.3 Concept Art & Storyboards - Second Iteration Prototype User Test Feedback Q1

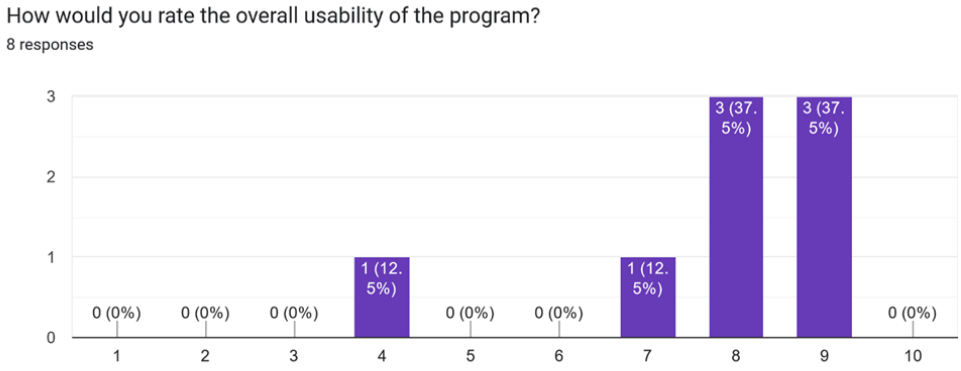


Figure 6.4 Concept Art & Storyboards - Second Iteration Prototype User Test Feedback Q3

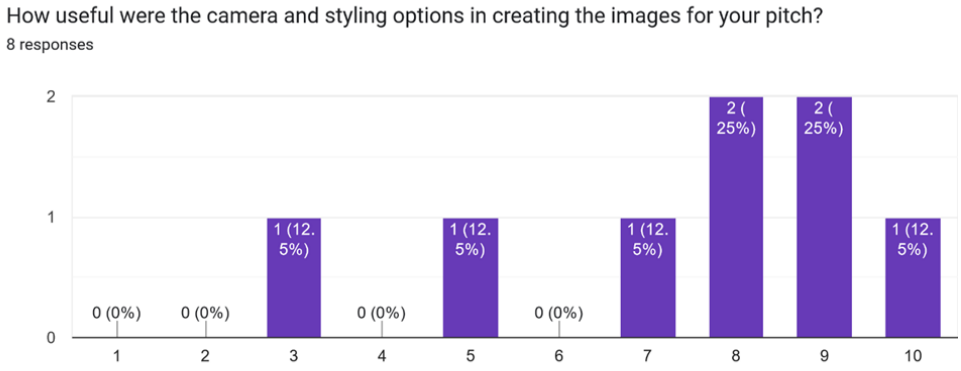


Figure 6.5 Concept Art & Storyboards - Second Iteration Prototype User Test Feedback Q4

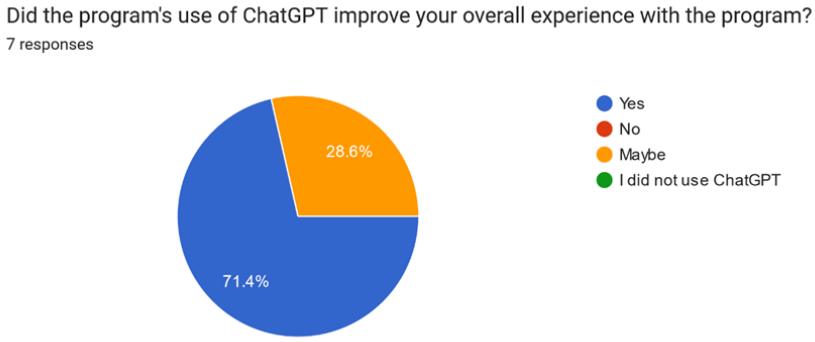


Figure 6.6 Concept Art & Storyboards - Second Iteration Prototype User Test Feedback Q12

User Tests Feedback Conclusion

Based on the feedback from user testing, several key points were identified. Overall, users found the web page navigation to be relatively easy. However, there were some bugs encountered during testing. One bug involved the selection of the "None" option, which still resulted in it being included in the user's input prompt or mixed with the ChatGPT refined prompts. Another issue was the lack of feedback to indicate the progress of image generation, such as a loading bar or visual indication. Furthermore, there was a need for improved handling of errors returned from the back-end server, especially when user prompts failed to pass the safety system of DALL-E 2.
Users also expressed a desire for the functionality of the "Variation" and "Regenerate" buttons, although it was understood that these features were not yet implemented due to time constraints. These issues have been acknowledged as the primary focus for the next iteration of the prototype, and efforts have been made to address and resolve them promptly.

6.1.3 Third Prototype

In the third prototype, there are few changes compared to the second prototype. Notably, the "Variation" and "Regenerate" buttons now have full functionality and perform as intended. The saved images are displayed below the generated images, and users have the option to remove a saved image if desired. Additionally, each saved image is accompanied by a caption text box where users can input their own captions if they choose to do so.

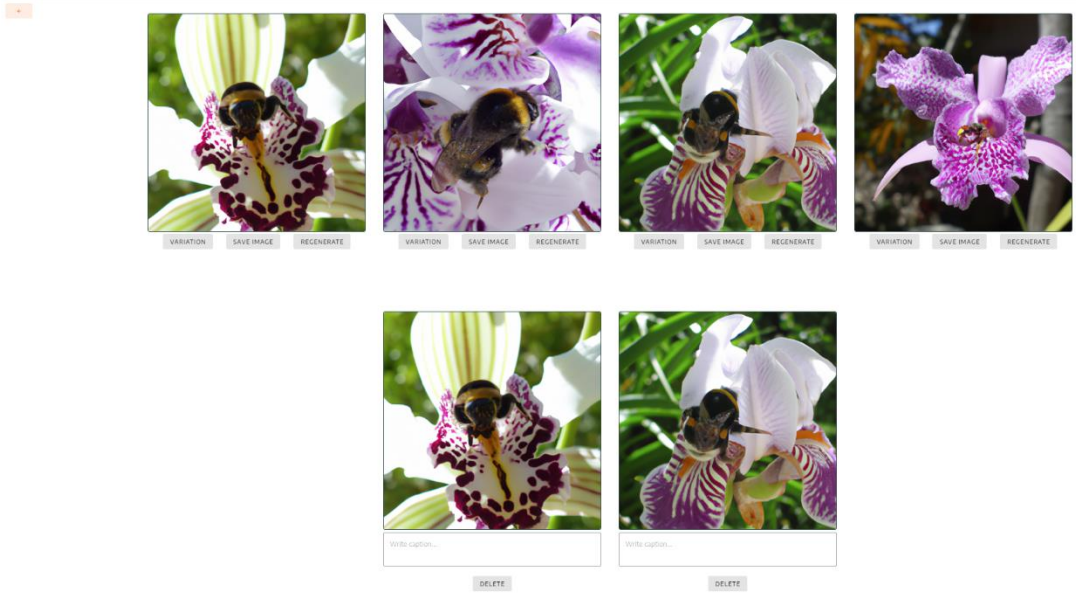


Figure 6.7 Third Iteration of Concept Art & Storyboards Prototype

6.1.4 User Interactions

Prompt Input

The prompt input area consists of two text box elements. The first text box allows the user to enter their own prompt, while the second text box (greyed out) is not editable, as it serves the purpose of displaying the ChatGPT-refined user prompt in case the user decides to ask ChatGPT for refining the prompt.

To generate images, the user can choose between using their own prompt or the ChatGPT-refined prompt, simply by clicking on the respective text box's "Generate" button.

It is important to note that in the background, the actual prompt undergoes processing to incorporate the selected options. This involves concatenating these options as a string to the original user prompt. However, if the user requests ChatGPT to refine the prompt, ChatGPT will utilize the selected options to generate a clear and more descriptive prompt, rather than simply concatenating them.

Figure 6.8 Concept Art & Storyboards Prototype - Prompt Input Area

Styling Options

The styling options in the prototype are user-friendly and easy to use. Users can adjust settings, for the example the "Camera proximity" by clicking on the option and selecting from a drop-down menu. The prototype allows for easy addition of new styling options if required, but currently, a limited number of options are available to maintain simplicity and showcase the approach's consistent styling capabilities.

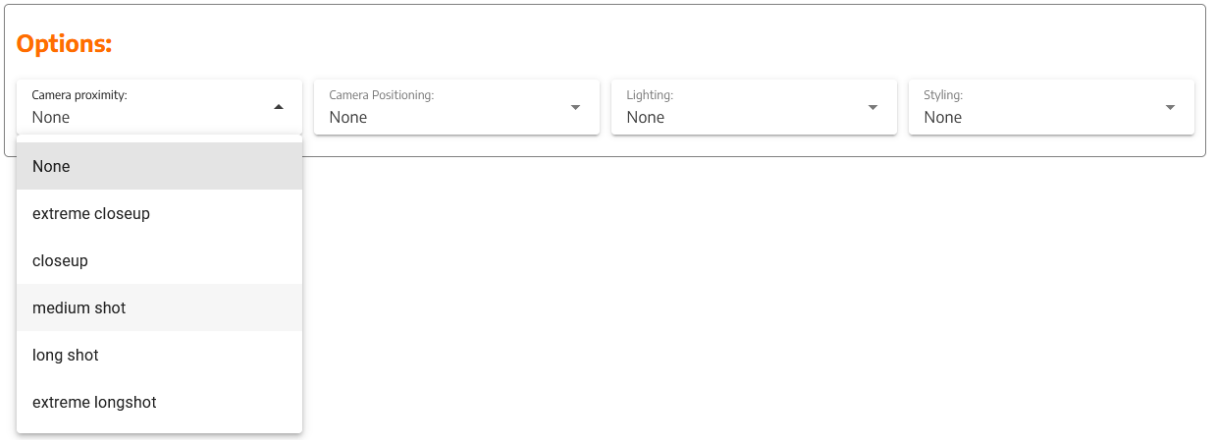


Figure 6.9 Concept Art & Storyboards Prototype - Styling Options

Image Regeneration & Variation

After generating the images, users are provided with options for creating variations or regenerating them. These options allow users to easily generate similar variations of the current image or initiate a complete regeneration with just one click. This functionality offers flexibility and convenience in exploring different iterations of the generated content.

Figure 6.10 Concept Art & Storyboards Prototype - Generated Image Options

Image Variation Example

In the following two images, an example of the "Variation" button functionality is shown. The Image "A" is the first generated image, while the image "B" shows a variation of the image "A", which was created by simply pressing the "Variation" button on the original image. As for now, the original image is lost and replaced with the new variation.

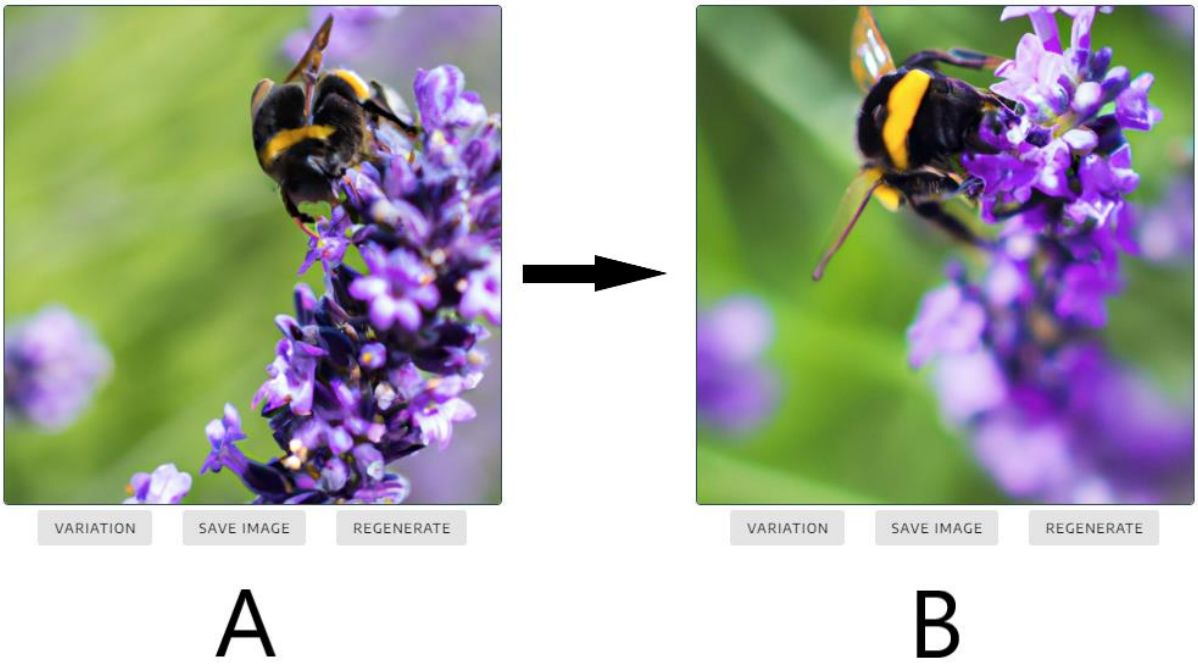


Figure 6.11 Concept Art & Storyboards Prototype - Image Variation Example

Saved Images

The saved images are displayed beneath the generated images. Currently, they do not possess any specific functionality, except for the ability to persist even when new images are generated. Below the saved images, a text box is provided for users to enter captions if desired.

To remove an image from the saved images section, users can press the "Delete" button.

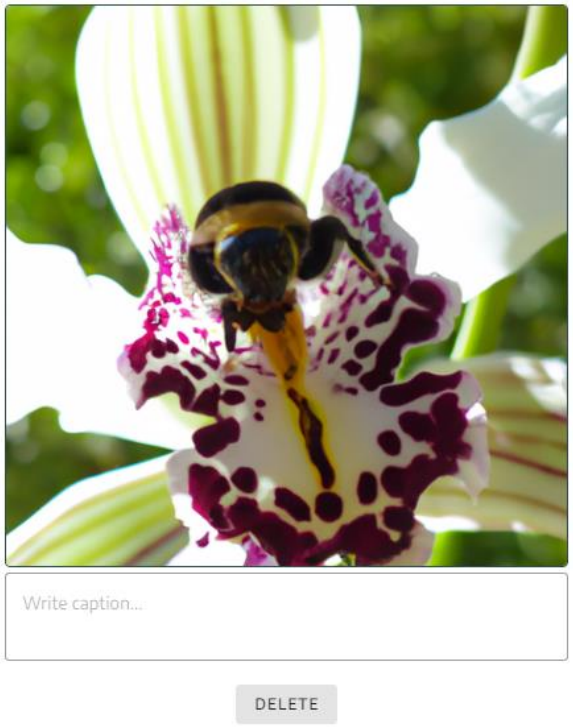


Figure 6.12 Concept Art & Storyboards Prototype - Saved Images

6.2 Videos & Animations

The focus of the user interface for the Text-To-Image tool is to provide a clear and intuitive design that is easy to comprehend and navigate. The goal is to ensure that users can interact with the tool effortlessly, while keeping the prompt input options simple and straightforward.

6.2.1 First Prototype

The initial prototype of the text-to-video tool encompasses all the intended features. It includes an animation duration slider, a box for editing the length of each animation prompt in seconds (or number of frames), an animation prompt text input, and settings related to 2D camera movements, such as zoom, pan in both directions, and rotation.
Furthermore, the user has the flexibility to set the seed manually or leave it blank to allow for a randomly selected seed.
These options provide users with control over the animation generation process and allow for customization based on their preferences.

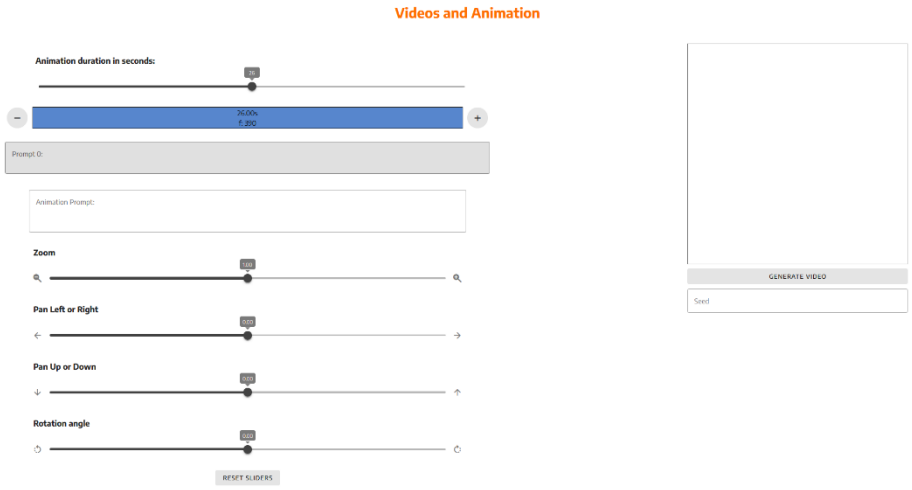


Figure 6.13 First Iteration of Video & Animation Prototype

User Tests

During the user testing of this tool, the same testing setup as the previously tested "Concept Art & Storyboards" tool was used.
In this user test, seven participants, colleagues, or peers of the graduation team members, took part in the testing process.

User Test Feedback Summary

In line with the previous tool prototype, this report will focus on presenting the technical feedback received during the user testing. The complete user feedback forms and responses will be included in the Graduation Dossier for reference.
Mentionable Feedback:

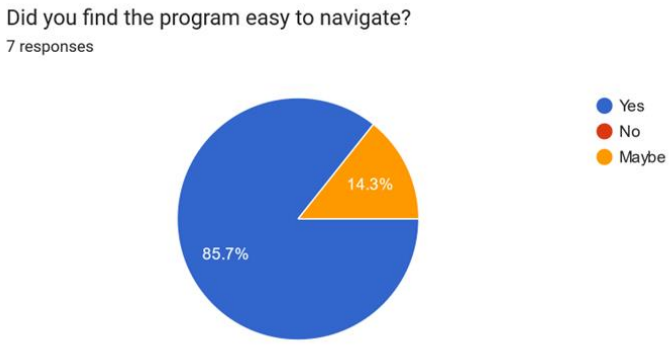


Figure 6.14 Video & Animation - First Iteration Prototype User Test Feedback Q4

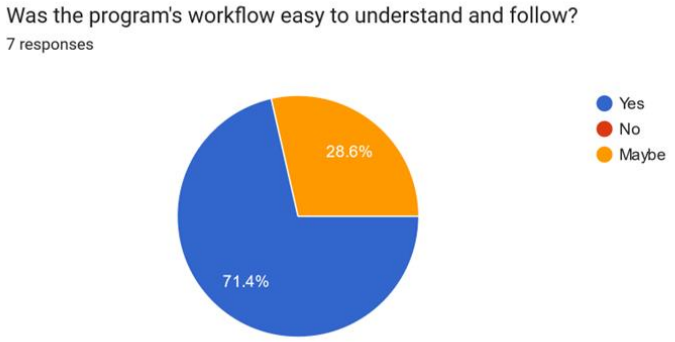


Figure 6.15 Video & Animation - First Iteration Prototype User Test Feedback Q5

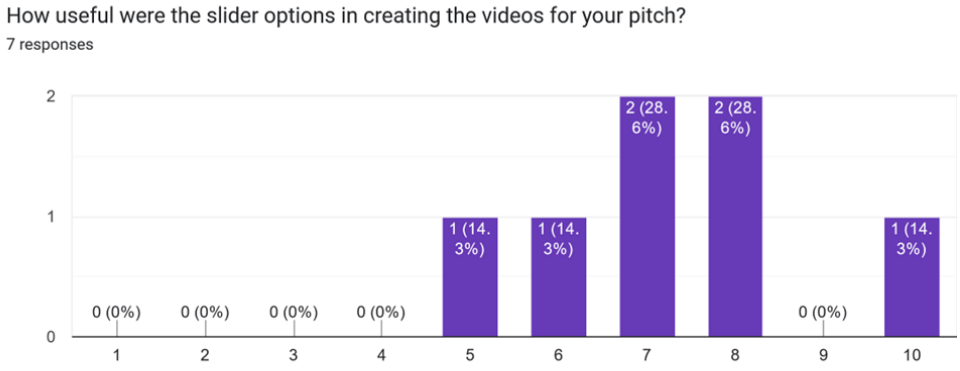


Figure 6.16 Video & Animation - First Iteration Prototype User Test Feedback Q7

User Test Feedback Conclusion

Based on the feedback received from the user test sessions, it is evident that the prototype of the text-to-video tool has a relatively low number of bugs and errors. However, there are a few areas that require improvement.

One key issue highlighted by the users is the lack of clear error handling on the front-end when errors are thrown from the back-end. Improving the error handling mechanism will enable better feedback to users when errors occur, enhancing the overall user experience and reducing confusion.

Another important point raised in the feedback is the users' request for more tooltips on the various elements of the prompt options. Users reported confusion while using the application, and they believe that additional tooltips would provide clear explanations and guidance on how to effectively use the different features.

In the next iteration of the prototype, the team will prioritize addressing these issues by enhancing error handling and implementing additional tooltips. These improvements will contribute to a smoother user experience and ensure that users can utilize the tool more effectively.

6.2.2 Second Prototype

In the second prototype, several new features have been introduced to enhance and guide the user through the animation prompting process. These features include tooltips that provide explanations for each UI element, ensuring clarity and assistance in navigating the tool effectively.
Additionally, below the video player, an overview of the current existing animation prompts is displayed. This feature aims to provide users with a convenient and comprehensive view of their prompts, allowing them to easily reference or view their prompts in one place.

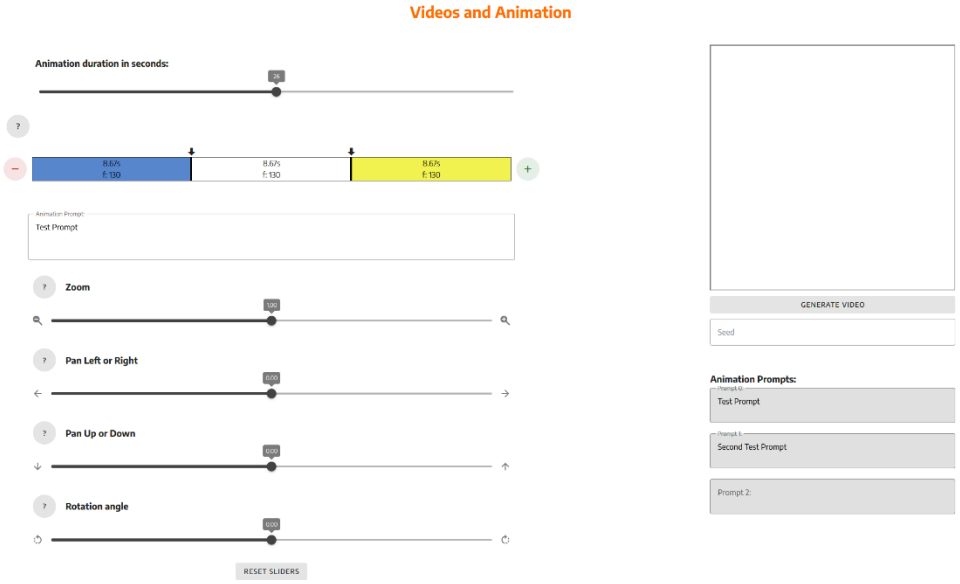


Figure 6.17 Second Iteration of Video & Animation Prototype

6.2.3 User Interactions

Animation Duration Slider

As previously mentioned, the animation/video duration slider serves the purpose of setting the duration of the final generated video in seconds. In the original version of the text-to-video tool, the duration was determined based on the user-provided "fps" and the total number of frames, which often led to confusion as users had to manually calculate the video duration by dividing the “max_frames” number by the given “fps” number.
Furthermore, the slider offers the advantage of clamping the minimum and maximum duration for generated videos. This addresses another issue present in the original tool, where the user was required to set a minimum value of “100” for the "max_frames" setting, otherwise it would not be possible to generate the video. The inclusion of the duration slider eliminates the possibility of such errors occurring, providing a more streamlined and error-free experience for users.

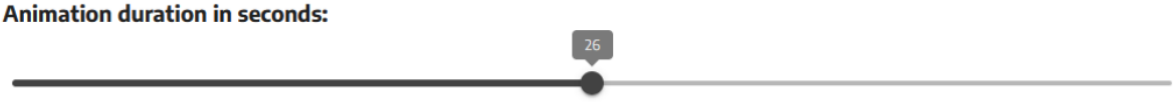


Figure 6.18 Video & Animation Prototype - Animation Duration

Animation Prompt Selector

The custom component allows users to easily modify the duration of each animation prompt individually. Users can add new prompts or remove existing ones using the designated (-) and (+)

buttons. Each animation prompt is displayed as a box, showing its duration in both seconds and frame count. Adjusting the duration is as simple as clicking and dragging the dividers between prompts. Visual cues, such as highlighting the selected prompt and indicating unfilled prompts, aid in usability. This component streamlines the management and adjustment of prompt durations for a smoother user experience.

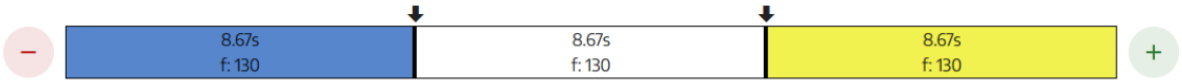


Figure 6.19 Video & Animation Prototype - Animation Prompt Selector

Animation Prompt Input & Camera Settings

When a box in the Animation Prompt Selector is clicked, the corresponding prompt text input and camera settings are displayed. Camera settings are represented as sliders with predefined value ranges. Icons on both sides of each slider visually demonstrate the effect of moving the slider towards a specific icon.

For example, the zoom slider has values below “1.00” for zooming out and values above “1.00” for zooming in. While the numerical representation may initially be confusing, the icons provide clear visual cues for the zoom behaviour at different slider positions.

Moreover, each slider option includes a "?" icon that, when hovered over, displays a hint tool bar with additional information and explanations about the slider's purpose and function. These features aim to enhance user comprehension and ease of use when adjusting camera settings.

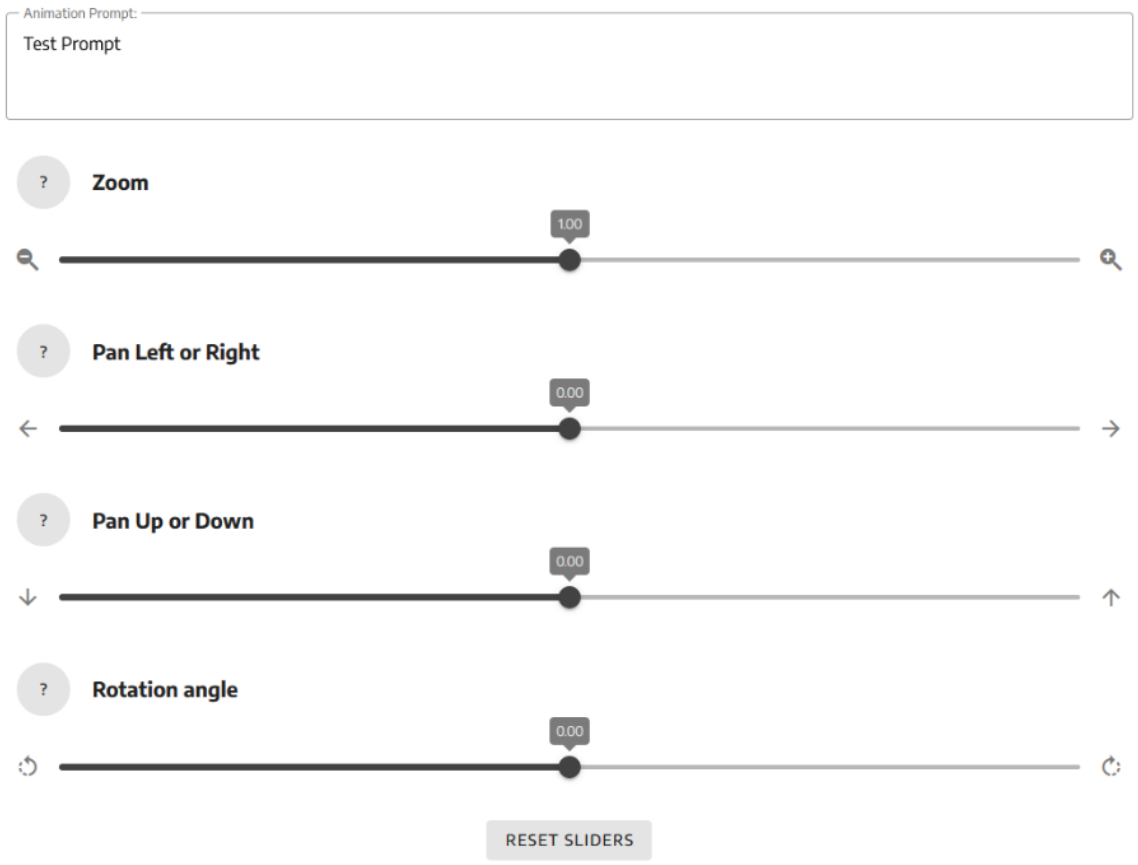


Figure 6.20 Video & Animation Prototype - Animation Prompt Input & Camera Settings

Animation/Video Generation & Seed selection

On the right side of the page, there are three elements. Firstly, a button initiates the video generation process. Adjacent to it, an input field allows users to manually enter a specific seed for video generation, otherwise the seed is randomly set if the input field is left empty. Lastly, an overview section displays the user's text input for each animation prompt. This section provides a reference point for reviewing and adjusting the input before generating the video.

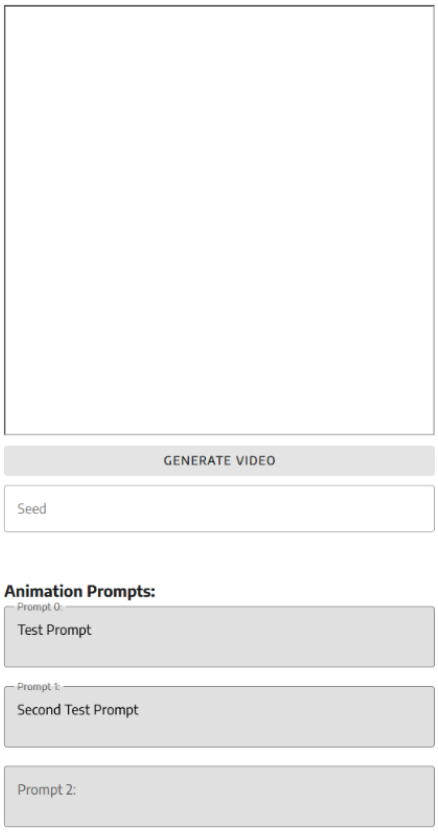


Figure 6.21 Video & Animation Prototype - Seed selection & Video Generation

6.3 Character Creation

The prototype for the character creation tool shares many similarities with the original D-ID text-to-character tool in terms of its features. It allows users to create characters based on given prompts and generate character dialogue videos.

6.3.1 First Prototype

The initial prototype contains the fully functional part for creating characters, which can be selected for generating the final character dialogue video. The right side shows the dialogue text input prompt and the "Generate Video" button for generating the video. As for now, the voice language, actor and styling are not available, and that is planned to be implemented in the next iteration of the prototype. For now, the language is set to English by default, and the voice actor and style is set to a default (woman voice with default voice styling).
The characters are generated using the provided prompts through an API call to the Stability AI (Stable Diffusion) Image Generator API.

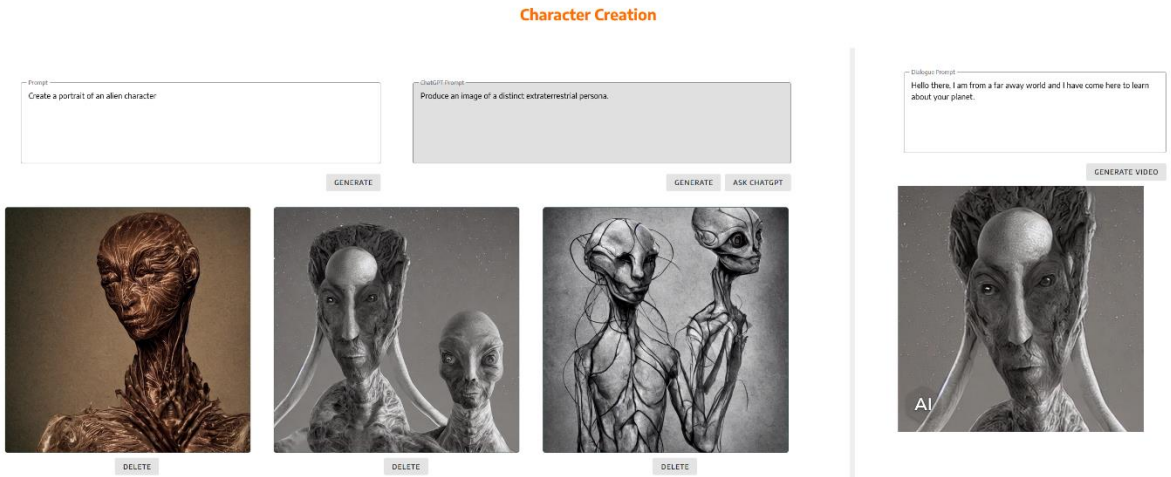


Figure 6.22 First Iteration of Character Creation Prototype

User Tests

Due to time constraints, user testing for the initial prototype was not conducted. Instead, internal testing was performed by the team's engineer and designer to assess usability and identify any errors or bugs. This testing confirmed that the prototype's technical aspects were sound, as pre-existing components had already undergone improvements. Although user testing provides broader insights, internal testing served as an initial evaluation to ensure functionality and identify immediate issues.

6.3.2 Second Prototype

The second prototype introduces additional features to enhance user customization. Users can now choose the language, voice actor, and voice styling options for their generated characters. The language selection includes English and Dutch, each with its own set of voice actors. The voice actors offer various voice styling choices. It's important to note that the prototype uses existing voices from the Azure AI Voice Gallery. Although the currently available options are limited in number, they effectively demonstrate the prototype's capabilities.

Furthermore, the selected character from the character creation section is now highlighted in orange, improving user experience, and facilitating quick identification of the chosen character.

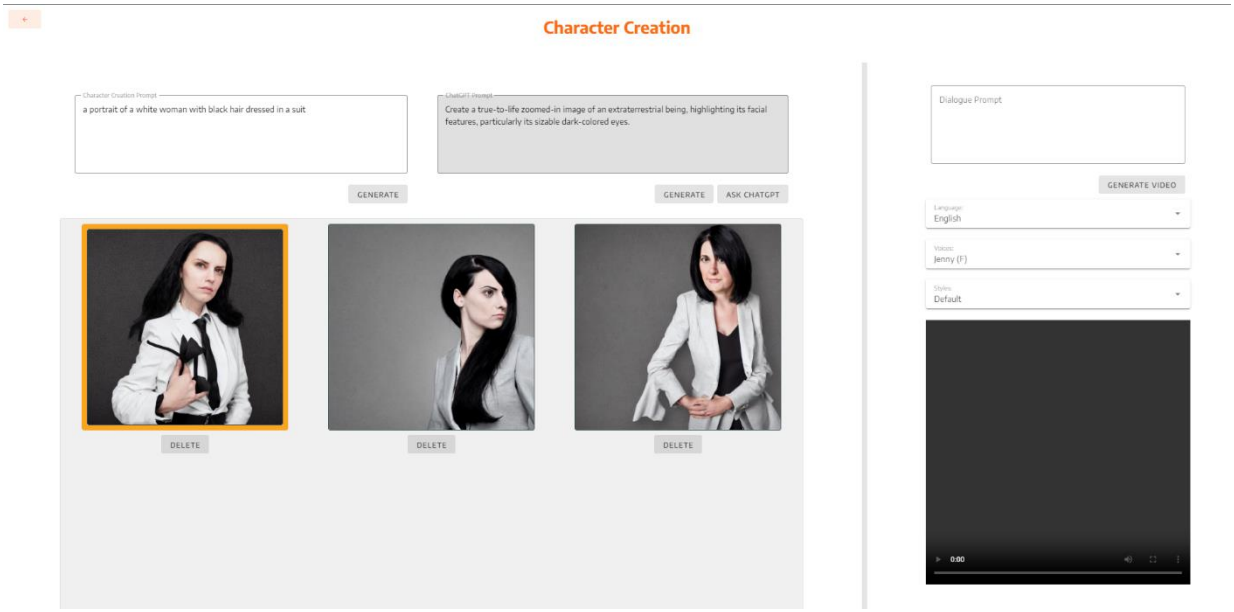


Figure 6.23 Second Iteration of Character Creation Prototype

To streamline the testing process and optimize resource utilization, the latest iteration of this prototype will be included in the final testing phase along with the other two models: "Concept Art & Storyboard" and "Videos and Animations." Combining the testing phase for all three models allows for a comprehensive evaluation of the overall system, assessing integration and interaction between the different models.

6.3.3 User Interactions

Character Generation Prompt Input

The text input boxes provided for character generation prompt follow the same functionality as those in the text-to-image model (Concept Art & Storyboards). The left box allows the user to input their prompt for the character they want to generate/create, while the right box is dedicated to show the refined user prompt by ChatGPT.

Figure 6.24 Character Creation Prototype - Character Creation Prompt Input

Character Selection

Users can easily select their desired character by left clicking on the character image. When a character is selected, it is visually highlighted with an orange border around the image. This highlighting feature allows users to quickly identify and keep track of their chosen character for generating the character dialogue video.

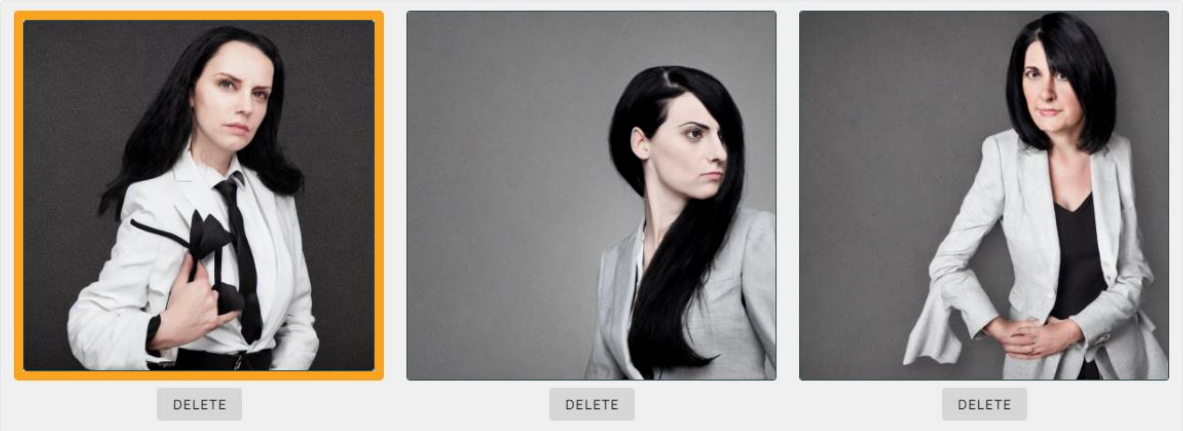


Figure 6.25 Character Creation Prototype - Character Selection

Dialogue Prompt & Voice Options

The dialogue prompt box enables users to enter the desired text for their character's dialogue. Below the prompt box, options are provided to select the language, voice actor, and voice styles to customize the character's speech, allowing for personalized voice and delivery preferences.

Additionally, a video player is available for previewing the generated character video, enabling users to view the character dialogue video once it has been generated.

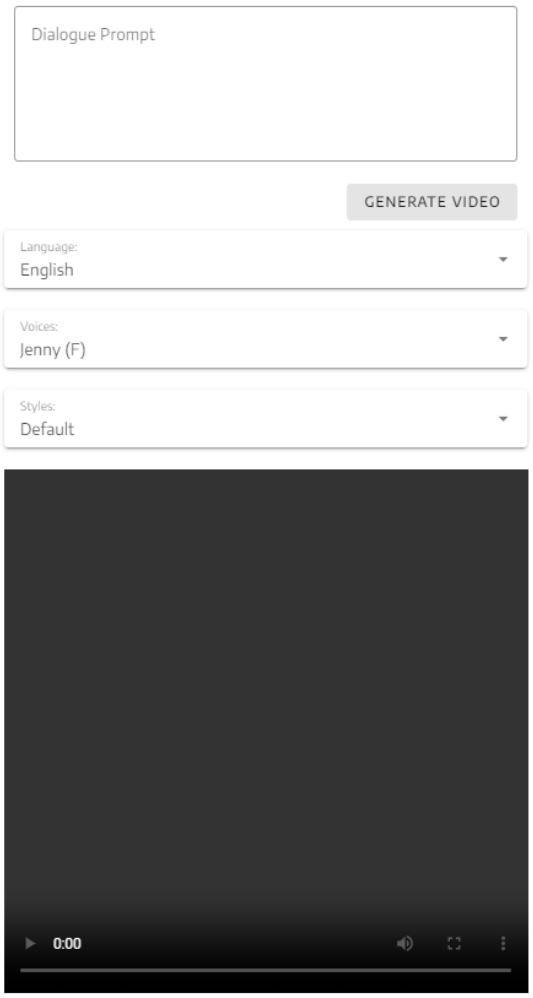


Figure 6.26 Character Creation Prototype - Dialogue prompt input & Language and voice options

7 Final test

The final testing phase follows the same setup as previous tests for each tool in the prototype. It incorporates various testing methods, including usability testing, observation, and questionnaires. The usability testing was done by the users, whom are colleagues and peers of this graduation team members. The observation method of testing was applied by the team members themselves by being always present during the user tests and observing the interaction of the users with the prototype and gaining insights. Nonetheless to support on any related questions that the users may have had or any bugs that may have occurred during testing.

In this final test, fewer user tests were conducted but more time was allocated to each test since each user tested all three tools. To get the most out of the test, gaining insight on how they see the three types of tools, which one they find more interesting and useful, would prove beneficiary to the overall results and feedback.
All the user testers filled a questionnaire at the end of each use test, and a summary of each question is shown below in graph or pie chart images.

In this test, all feedback, encompassing both technical and non-technical aspects, is displayed. This comprehensive approach allows for a holistic evaluation of the web application prototype, considering its overall usability and technical features. Filtering the technical feedback from all the given feedback is not deemed necessary in this case.

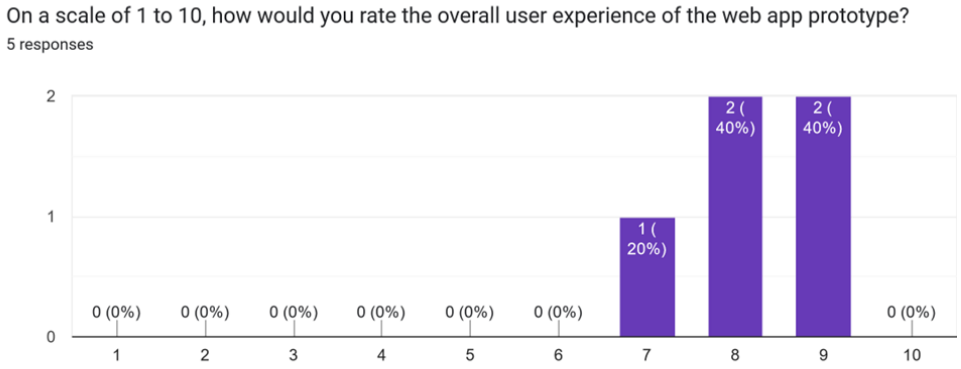


Figure 7.1 Script Forge Prototype - Final Test - Q1

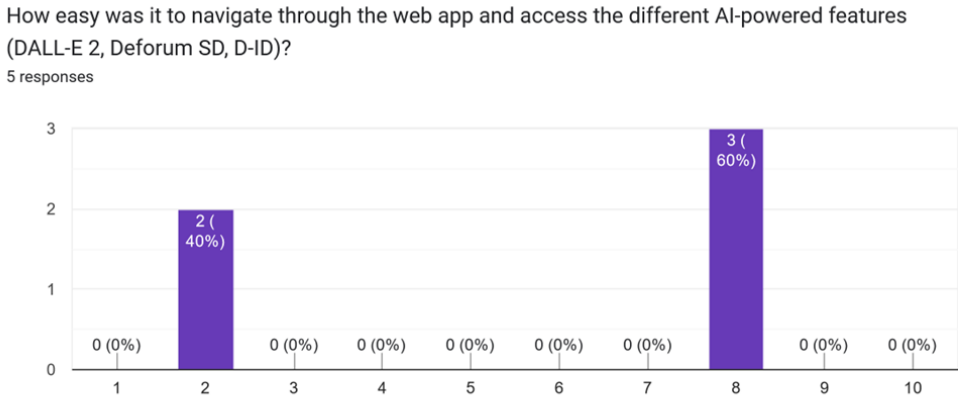


Figure 7.2 Script Forge Prototype - Final Test - Q2

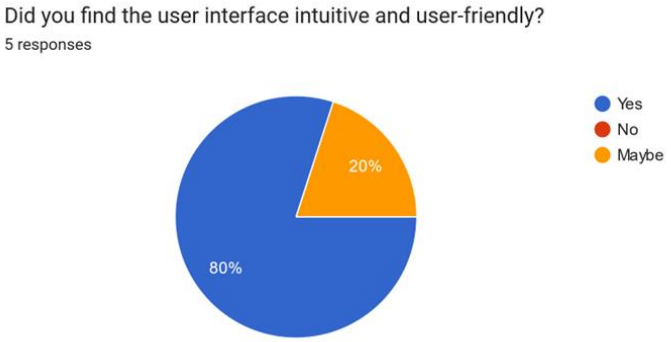


Figure 7.3 Script Forge Prototype - Final Test - Q3

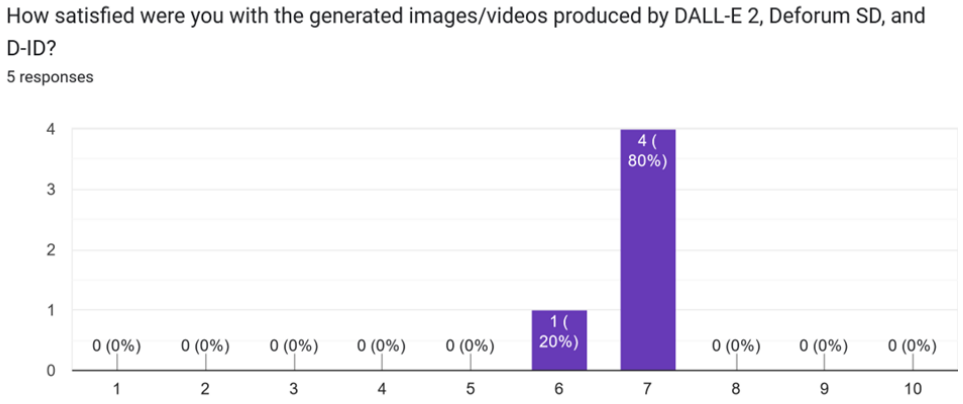


Figure 7.4 Script Forge Prototype - Final Test - Q4

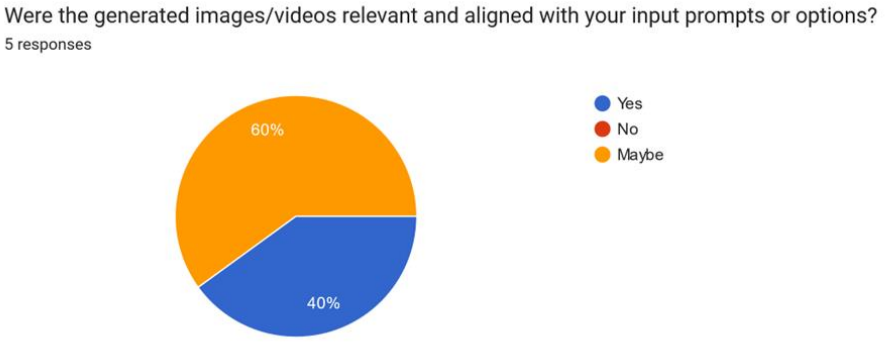


Figure 7.5 Script Forge Prototype - Final Test - Q5

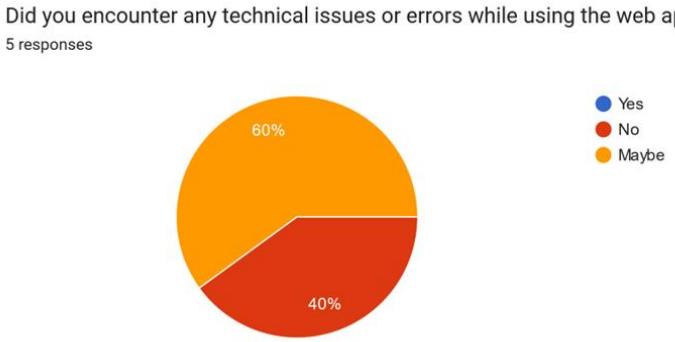


Figure 7.6 Script Forge Prototype - Final Test - Q6

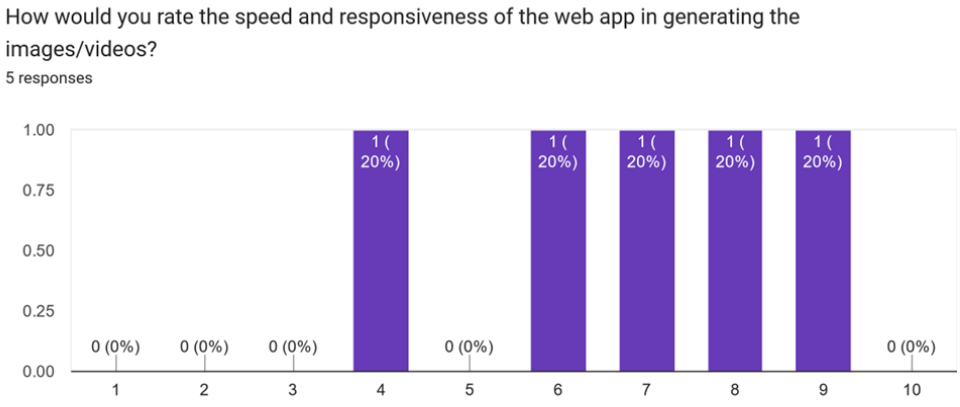


Figure 7.7 Script Forge Prototype - Final Test - Q7

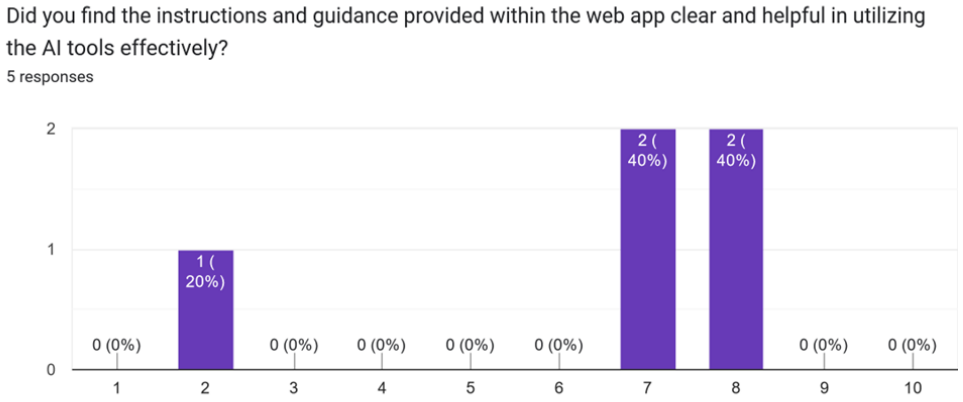


Figure 7.8 Script Forge Prototype - Final Test - Q8

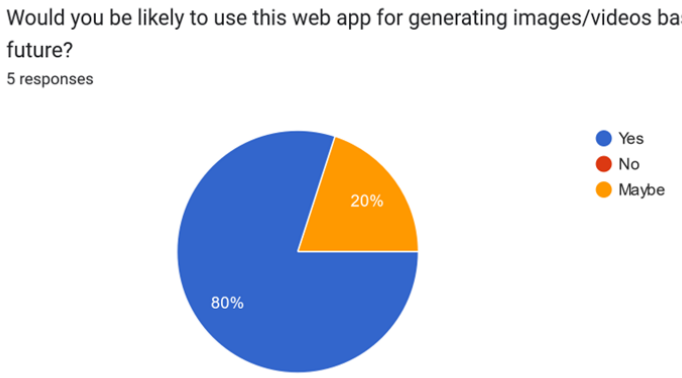


Figure 7.9 Script Forge Prototype - Final Test - Q9

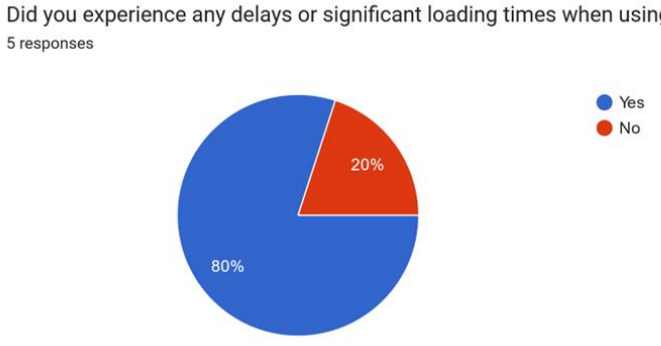


Figure 7.10 Script Forge Prototype - Final Test - Q10

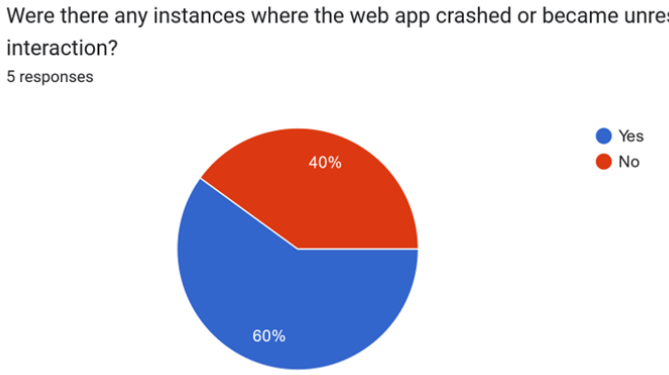


Figure 7.11 Script Forge Prototype - Final Test - Q11

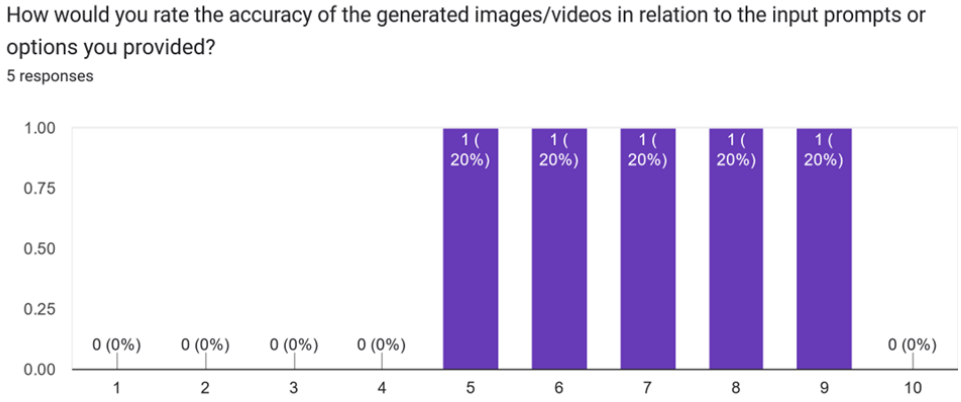


Figure 7.12 Script Forge Prototype - Final Test - Q12

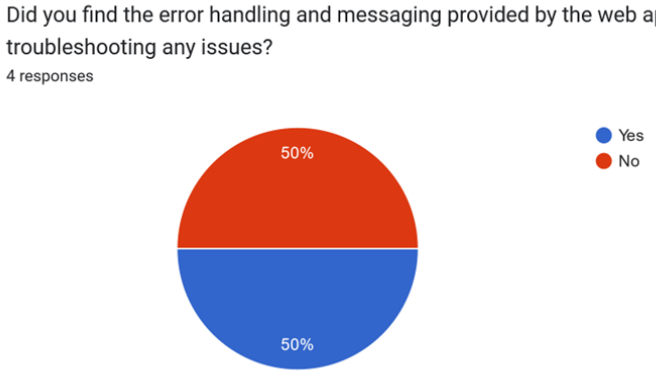


Figure 7.13 Script Forge Prototype - Final Test - Q13

8 Results

8.1 User Tests Results

Usability
Many of the user testers rated the overall user experience as great, although some inconveniences were encountered in certain parts of the application.

User Interface
Approximately 80% of the testers consider the user interface to be intuitive, user-friendly, and somewhat easy to navigate.
More than 70% of the testers found the provided instructions and guidance for using the tool somewhat helpful.

Generated Content
Most of the user testers were left somewhat satisfied with the generated content based on their input and available options, although the responses vary.

Functionality
Only half of the user testers found the error handling messages helpful.
More than half of the user testers reported occasional unresponsiveness in the application.
The speed and responsiveness of the web app received good ratings, although opinions were distributed across the rating scale.

8.2 Client Final Showcase

After showcasing the final prototype to the client at their office, the client expressed high satisfaction with the final prototype, acknowledging its status as a preliminary version. Despite the product being still in the prototype phase, the client highly appreciates the main concept idea and recognizes its potential as a valuable addition to the current pitch review system for NPO workers. The client firmly believes that with further development and improvements, the prototype can evolve into an exceptional tool. This positive feedback underscores the client's confidence in the project's direction and serves as strong motivation for the team to continue refining and expanding the prototype into a remarkable solution for the NPO workers' pitch review process.

9 Conclusions

After analyzing the results, from the user tests and the client showcase, it is evident that the final prototype effectively fulfills the initial requirements set by the client. The prototype boasts several notable features that contribute to its overall functionality. However, there are still areas that warrant further refinement and improvements in future iterations of the prototype.

In general, the user interface was found to be intuitive by most user testers. They encountered little to no difficulties while navigating through the AI tools within the application. This positive experience can be attributed to the improvements made based on feedback received from previous user tests conducted on these specific tools.

It was also noticed that users' proficiency in creating clear and coherent prompts was related to their language expertise. Users with a strong command of the language were generally better at formulating prompts that yielded the better results. This reinforces the importance of considering language skills when addressing the challenges faced during prompt creation.

The impact of users' language skills was evident in the user test feedback, as some participants rated their experience mediocre at best, due to challenges in prompt formulation. It became clear that language proficiency played a significant role in achieving satisfactory results. However, the inclusion of ChatGPT prompt refinement proved to be a valuable addition to the solution, helping users improve their prompts and enhance the overall experience.

Some of the prototype's limitations can be attributed to the engineer's lack of expertise in web development. This resulted in significant time being dedicated to studying and researching various web development technologies, including both back-end and front-end development. The learning process surrounding web development has been crucial in identifying areas for improvement and expanding the engineer's skill set.

10 Recommendations

Despite the final prototype successfully meeting the client's requirements, there is still room for improvement. Based on the current results and identified shortcomings, the following recommendations are suggested to enhance the application:

- Implement state handling: Ensure that the end user's progress and generated content are saved and retained when seamlessly switching between different tools within the application.
- Implement a prompt history feature: Allow users to access and reuse previous prompts if needed, providing convenience and efficiency in generating content.
- Enable cloud storage for generated content: Integrate a feature that allows end users to save their generated content in the cloud, ensuring centralized access for NPO workers and facilitating easier collaboration.
- Implement safety guidelines: Incorporate safety features and guidelines to prevent errors caused by inappropriate or problematic input prompts, ensuring a smoother user experience.
- Provide better explanation of features: Enhance user understanding by incorporating tooltips and tutorials that offer clear explanations of the application's functionalities.
- Integrate image & video editing tools: Expand the application's capabilities by integrating image and video editing tools, allowing users to modify and enhance the generated content.

By implementing these recommendations, the application can further enhance its functionality, user experience, and overall effectiveness.

11 Reflection

During my graduation internship as a game developer, I embarked on a journey to learn full stack web development from scratch. It was a challenging but rewarding experience that has significantly enhanced my skill set. One tool that played a crucial role in my learning and development journey was ChatGPT. It served as a valuable resource, providing me with insights and saving me precious time as I delved into the world of web development.

In addition to mastering web development technologies such as HTML, CSS, Python, Flask, Vue.js, JavaScript, Axios, and more, I had the opportunity to explore the realm of generative AI models. This exposure allowed me to witness the vast potential of these models and opened my eyes to the possibilities they offer in various domains.

A pivotal aspect of my internship was the opportunity to work on projects that would be utilized by a diverse user base. This exposure enabled me to grasp the intricacies of the design process, including iteration, prototyping, and user testing. Through this hands-on experience, I honed my problem-solving skills and gained insights into how users perceive and interact with the products I developed.

Overall, my time spent during the five-month internship has been instrumental in expanding my knowledge and skills in web development and generative AI. I am proud of the progress I have made and the practical experience I have gained in creating solutions that cater to the needs of users. This experience has shaped me into a more well-rounded developer, equipped with a comprehensive understanding of the development lifecycle and the importance of user-centric design.

12 References

About NPO. (n.d.). Over NPO. https://over.npo.nl/organisatie/about-npo

Our mission. (n.d.). Over NPO. https://over.npo.nl/organisatie/about-npo/our-mission

Home - maatschappelijke waarde. (n.d.). NPO Maatschappelijke Waarde. https://maatschappelijkewaarde.npo.nl/

Wikipedia-bijdragers. (2023). Stichting Nederlandse Publieke Omroep. *Wikipedia*. https://nl.wikipedia.org/wiki/Stichting_Nederlandse_Publieke_Omroep

The DALL·E 2 prompt book. (n.d.-a). https://pitch.com/v/DALL-E-prompt-book-v1-tmd33y

Andrew. (2023, April 23). How to make a video with Stable Diffusion (Deforum) - Stable Diffusion Art. *Stable Diffusion Art*. https://stable-diffusion-art.com/deforum/

Flask Vs Django: Which Python Framework to Choose? (2022, July 29). InterviewBit. https://www.interviewbit.com/blog/flask-vs-django/

Python and Flask Tutorial in Visual Studio Code. (2021, November 3). https://code.visualstudio.com/docs/python/tutorial-flask

OpenAI API. (n.d.-c). https://platform.openai.com/docs/guides/rate-limits/overview

OpenAI API. (n.d.-b). https://platform.openai.com/docs

Vue.js - The Progressive JavaScript Framework | Vue.js. (n.d.). https://vuejs.org/

Welcome to Flask — Flask Documentation (2.3.x). (n.d.). https://flask.palletsprojects.com/en/2.3.x/