Software Engineering Department
ORT Braude College

23-2-D-19

VidWizard: Generate Smart Video Content Using AI Tools

Supervisors:

Alexander Keselman

Authors:

Shaked Arish: 311420384
Sean Friedman: 204404511

Table of Contents

1. Abstract

In today's digital age, the ability to create captivating and interesting content is crucial, especially on popular social media platforms like Twitter, Facebook, and Instagram. Whether it's for marketing, education, or entertainment purposes, social media become a powerful tool for individuals, businesses, and organizations in order to succeed in the digital world.
The rapid advancement of artificial intelligence (AI) technology has opened up new possibilities for content creation. In this project, we present an innovative solution that leverages the power of AI tools to generate smart content, particularly short videos, and integrate it into a user-friendly and accessible platform. Through the integration of multiple AI technologies and APIs, our platform streamlines the content creation process, eliminating the need for extensive manual production and technical skills. Users can simply input their preferences, topic, and desired visual elements, and our platform will take care of the rest. From generating customized scripts to synthesizing voiceovers and incorporating visual elements, our platform ensures a tailored and engaging video creation experience.
We address the limitations of existing approaches by offering a unique combination of automation and customization. Users have the freedom to personalize and edit the AI generated content throughout the process, making the final video truly unique and aligned with their vision.
Our goal is to make video content generation accessible to a wider audience, regardless of their video editing skills and technical expertise. empowering them to share their ideas and stories effectively through the help of AI powered technologies.

2. Introduction

The effect of high-quality content in social media becomes crucial for businesses and individual people alike, in order to attract and engage audiences and ultimately boost their exposure. There are several approaches to generating smart content, each with varying levels of effectiveness and user-friendliness. In this introduction, we will explore two common methods: manual video production and template-based automated systems. These methods have their advantages and limitations, highlighting the need for a more innovative solution.

2.1 Motivation

Manual video production has been the traditional approach for content creators, offering full creative control but requiring extensive time, resources, and technical expertise. On the other hand, template-based automated systems provide a simplified process but often result in generic and repetitive videos that fail to meet user preferences. These limitations call for a transformative solution that combines the power of AI with a user-friendly interface.

2.2 Overview

Our project aims to change the process of generating smart content by developing a platform that harnesses the power of AI while maintaining creative control and ease of use.
Through our platform, by answering a straightforward form, users can easily express their content topic, preferences, and requirements. The information provided by the user through the form will serve as the foundation for generating the video, by integrating the user input with a precise, pre-built template that leverages advanced AI tools, our system generates a script of the intended video, offering the users selection of voices for voiceover and find relevant visuals for it.
This innovative approach bridges the gap between technical expertise and content creation, enabling individuals from all backgrounds to produce high-quality video content with ease.

In the following paper, we will further elaborate on the idea of our project, his expected work process, the technologies involve in it, and the expected challenges.

3. Background & Related Work

3.1 Artificial Intelligence

"Artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience." [1]

Artificial intelligence (AI) describes the development of computer systems that are capable of handling tasks that traditionally require human intelligence. Computer science, mathematics, psychology, neuroscience, and other related fields are all combined in AI.
AI seeks to develop intelligent machines with human-like capacities for perception, learning, and problem-solving. The objective is to create systems that can comprehend and analyze natural language, spot trends, make judgments, learn, solve problems, and adapt themself.
Narrow AI and General AI are the two main subtypes of AI. Weak AI, also known as narrow AI, describes systems that are created to carry out functions inside specified domains, such as voice assistants and picture recognition systems. These AI systems are exceptional at what they are taught to do, but they are limited in their ability to generalize.
In comparison, general AI, usually referred to as strong AI, describes AI systems that can comprehend, learn, and apply knowledge across several domains and tasks, much like human intelligence. The goal of strong AI is to create machines that can reason and think like people.
There are various industries where AI can be used, including healthcare, banking, transportation, gaming, customer service, and education. AI has the ability to increase productivity and can play a significant role in the creation of cutting-edge services.

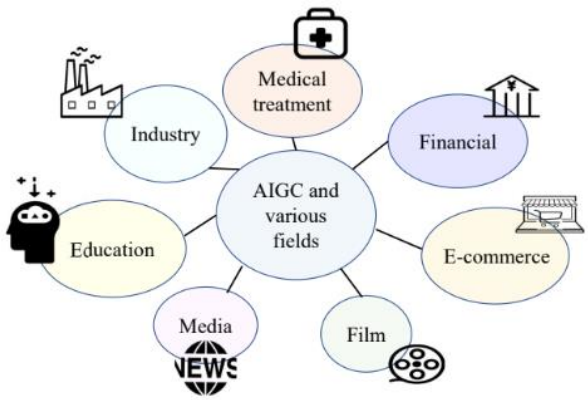

Figure 2: The combination of AI Generated Content in other fields

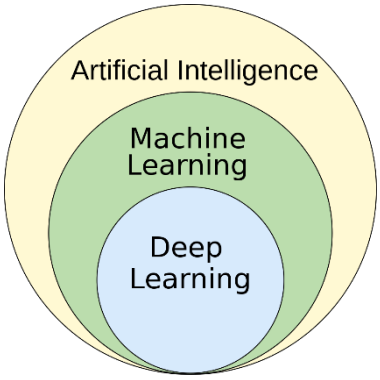

Figure 1: AI subfields

3.2 Machine Learning

“Machine learning (ML) is a field devoted to understanding and building methods that let machines "learn" – that is, methods that leverage data to improve computer performance on some set of tasks”.[2]

Machine learning is a subcategory of artificial intelligence that focuses on enabling computers to learn and make predictions and decisions without being programmed to do it. It is a field that includes developing algorithms and models that learn patterns and relationships from the data.
Machine learning can be separated into 4 main types:

- Supervised Learning
- Unsupervised Learning
- Semi Supervised Learning
- Reinforced learning

Supervised learning: The model is trained on labeled data.
Unsupervised learning: The model is trained on unlabeled data.
Semi-supervised learning: The model is trained on a mix of labeled and unlabeled data.
Reinforced learning: involves training an agent to communicate with an environment and learn optimal actions based on rewards or feedback.

The difference between supervised and unsupervised learning is that in unsupervised learning the output data is not given. The learning process occurs by using the relation and connection in the data. Moreover, unsupervised learning doesn't have training data.
Machine learning can be used in different applications, including image and speech recognition, natural language processing, recommendation systems, fraud detection, autonomous vehicles, and many more.

3.3 API

"An Application Programming Interface (API) is a way for two or more computer programs to communicate with each other. It is a type of software interface, offering a service to other pieces of software." [3]

APIs are common in the world of software development. They are the communication feature that enables software applications to interact and exchange data with each other. We can find APIs in use in web-based systems, operating systems, database systems, and even the hardware in our computers and smartphones. For example, whenever we sign into a social media app on our phone, it's an API that retrieves the login information from the server. For developers, APIs simplify the process of integrating features into their software. Instead of building functionality from the start, developers can simply

call the API and use the output results as they wish in their application. This makes the code more reusable.

3.3.1 AI API's

While APIs have been around for quite some time, AI APIs are a new concept. An AI API is an interface that allows developers to integrate AI capabilities into their applications. They leverage machine learning and artificial intelligence techniques to extract relevant data and interact with server software or other applications. According to a study made in 2021 by McKinsey, 56% of organizations have already adopted AI in at least one business function, often through the use of AI APIs. [4]

3.4 AI Tools

3.4.1 Text-To-Speech (TTS)

Text-to-Speech (TTS) converts written text into spoken words, new technologies in the field make use of deep learning and artificial intelligence among other advanced techniques to capture linguistic patterns and generate high quality, fluency, and natural sounding speech.
TTS systems usually offer the user a voice selection, allowing users to choose from a variety of voices that represent different genders, ages, and accents according to their needs and desire, resulting in a more personalized and engaging speech.
The applications of TTS technology are vast, it plays a crucial role in various domains, ranging from accessibility for individuals with visual impairments to multimedia content creation and language learning tools.

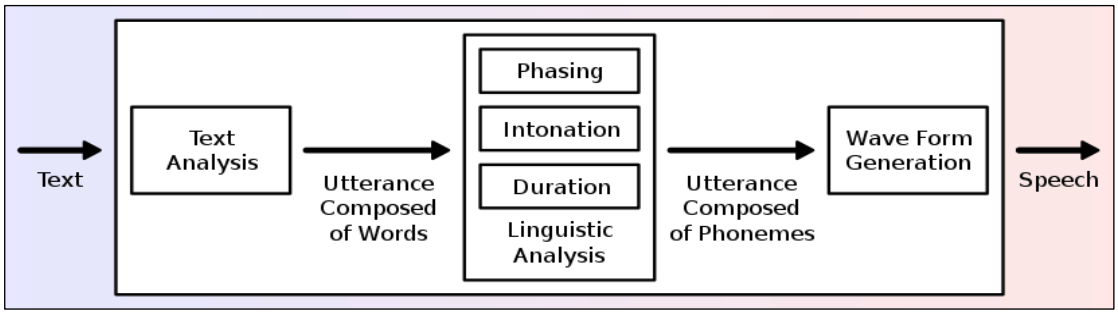

Figure 3: Overview of TTS System

3.4.2 Text-To-Video (TTV)

"A text-to-video (TTV) model is a machine learning model which takes as input a natural language description and produces a video matching that description."[5]

Text-to-video tools streamline the time-consuming aspects of video production, making it effortless for users to generate engaging video content and opening up a world of possibilities. While text-to-video model innovative technology has gained significant traction in recent years, thanks to advancements in AI algorithms and machine learning techniques, the field is still in its early stages, with ongoing research and advancements driving its evolution. Access to text-to-video generation technology is currently limited, and the quality of the generated videos may not always meet the expectations of the users.

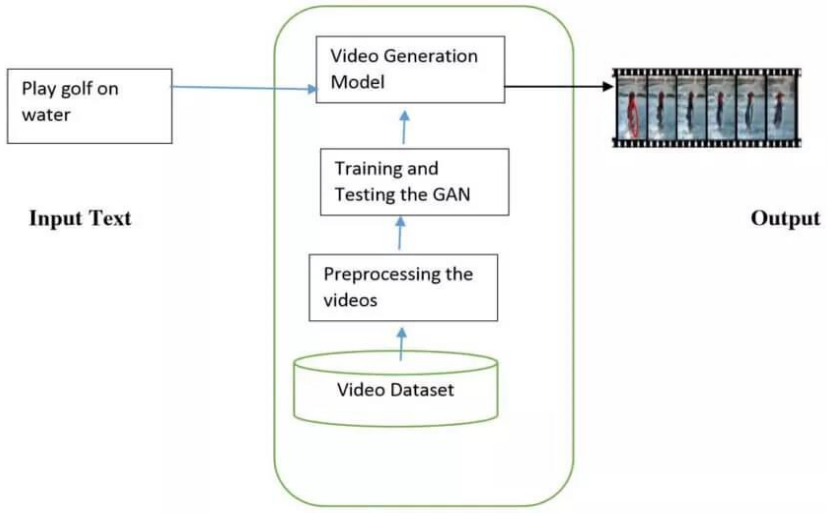

Figure 4: Architecture of Text-To-Video Generation System

3.4.3 ChatGPT

Is an advanced language model developed by OpenAI. As one of the most powerful language models available. It operates by utilizing deep learning algorithms to understand and process natural language inputs. It has been trained on an extensive array of text sources, including books, articles, websites, and other textual materials, allowing it to have a broad knowledge base on various subjects up until its knowledge cutoff date in September 2021.

His primary goal is to engage in interactive conversations with users, offering informative and coherent responses to their queries. It is designed to comprehend and generate human-like text, providing assistance, answering questions, and offering suggestions based on the input it receives. ChatGPT has made a remarkable impact on the world, particularly in the field of artificial intelligence and natural language processing.

One of the notable breakthroughs of ChatGPT is its ability to generate human-like text responses. It has learned to understand and generate coherent and contextually relevant language. This breakthrough has opened up new possibilities for human-like conversations and interactions with AI systems.

Furthermore, Open Ai's introduction of the GPT-3 API has allowed developers and businesses to integrate ChatGPT into their own applications and services seamlessly. This accessibility has empowered a wide range of industries and individuals to leverage the power of ChatGPT for their specific needs. The impact of ChatGPT goes beyond its technical advancements. It has sparked widespread interest and engagement from researchers and developers especially in the field of AI ethics and fairness.

3.5 AI – Ethics

"At the 2018 ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*), an associate professor at Princeton University gave a talk outlining 21 different definitions of fairness. This goes to show how nuanced and context-sensitive the concept of "fairness" is. its definition differs across cultures, across societies, and even over time."[6]

There is no single definition of fairness. When we're talking about AI systems, we need to think carefully about what fairness means in that context. If we don't carefully consider fairness throughout the entire process of building an AI system, it could unintentionally produce unequal outcomes for people from different gender, race, religion etc.

3.5.1 Bias In AI Systems

Bias can enter your AI system at different stages of our development, starting from the diversity of the team working on it and how they gather and classify data, to the specific dataset chosen and the goals set for the system.

<u>Unfairness main causes:</u>

- A training dataset that isn't representative - Machine learning models and AI systems learn to identify patterns in large datasets that given by humans. If those datasets are lacking data on certain groups, it will lead to less accurate results for those underrepresented groups.

- A training dataset that has social bias inside - A training dataset can contain social and historical data in it. One commonly referenced example of this is the gender bias learned through an AI approach called Natural Language Processing (NLP), NLP is designed to understand human language, but in order to learn human language it needs a training data in the size of all Wikipedia or Google News for example, this training data was created by humans, with all of our imperfections and bias written in. This becomes the text that the AI learns from as a result NLP models may cause a wrong connection between professions and genders.

3.5.2 Unfairness examples in AI systems

- In a paper by Joy Buolamwini and Timnit Gebru revealed that facial analysis algorithms misclassified darker-skinned females while achieving high accuracy for lighter-skinned males. The darker-skinned females saw error rates of up to **34.7%**, whereas the error rate among lighter-skinned males was **only 0.8%**. The training dataset didn't contain enough examples of darker-skinned females for the model to perform well for that group.[7]

- in 2018, Amazon's hiring software was criticized for unfairly downgrading job applications that mentioned the word "women's".[8]

- Apple's credit card algorithm showed a clear bias against women, this was evident when Steve Wozniak, the co-founder of Apple, discovered that his wife was given a credit limit that was 10 times lower than his, even though they shared all their financial resources and accounts. [9]

3.6 Pexels

Pexels was founded in 2014 by twin brothers Ingo and Bruno Joseph in Fuldabruck, Hesse. It's become a go-to platform for designers, content creators, and marketers and offers a vast collection of over 3.2 million free stock photos and videos. The website provides high-quality visual assets for personal and commercial use without the need for attribution. Pexels' library covers a wide range of categories, styles, and themes, ensuring suitability for various projects. All content on Pexels is licensed under the Creative Commons Zero (CC0) license, allowing free usage, modification, and distribution. It offers a user-friendly interface with easy browsing and searching capabilities. Pexels is a convenient and reliable resource for accessing visually appealing content without licensing fees or attribution requirements. Its commitment to free usage makes it valuable for individuals and businesses alike. In our project we can make a request to the Pexels free API and use its huge library of images and videos for our benefit.

3.7 Node.JS

Is an open-source, cross-platform, back-end JavaScript runtime environment that runs on the V8 engine and executes JavaScript code outside a web browser. Node.js lets developers use JavaScript to write command line tools and for server-side scripting and running scripts server-side to produce dynamic web page content before the page is sent to the user's web browser. Consequently, Node.js represents a "JavaScript everywhere" paradigm, unifying web-application development around a single programming language, rather than different languages for server-side and client-side scripts. Though .js is the standard filename extension for JavaScript code, the name "Node.js" doesn't refer to a particular file in this context and is merely the name of the

product. Node.js has an event-driven architecture capable of asynchronous I/O. These design choices aim to optimize throughput and scalability in web applications with many input/output operations, as well as for real-time Web applications (e.g., real-time communication programs and browser games). The Node.js distributed development project was previously governed by the Node.js Foundation and has now merged with the JS Foundation to form the OpenJS Foundation, which is facilitated by the Linux Foundation's Collaborative Projects program.

3.8 JSON

JavaScript Object Notation is an open standard file format and data interchange format that uses human-readable text to store and transmit data objects consisting of attribute value pairs and arrays (or other serializable values). It is a very common data format, with a diverse range of applications, one example being web applications that communicate with a server. JSON is a language-independent data format. It was derived from JavaScript, but many modern programming languages include code to generate and parse JSON-format data. JSON filenames use the extension .json.

3.9 Video Content & Social Media

3.9.1 Video Content

Refers to audiovisual media that is created and shared digitally. It encompasses a wide range of formats, including short videos, tutorials, promotional videos, vlogs, and more. Video content has gained immense popularity in recent years due to its engaging nature, ability to convey information effectively, and its compatibility with various platforms and devices.

3.9.2 Social Media

Social media platforms have become powerful tools for individuals, businesses, and organizations to express ideas, share information, build brand awareness, and connect with their target audience. In this crowded online landscape, creating interesting and relevant content is crucial to make an impact, gain trust, and foster meaningful interactions with followers.

3.9.3 Video Content on Social Media

Video content in particular has captured the attention of social media users due to its visual appeal and ability to convey information effectively. Short videos especially have gained popularity as they provide concise and easily consumable content that fits well within the fast-paced nature of social media like TikTok, YouTube and etc. [11]

3.10 Pictory.AI

Pictory is an AI tool developed by a team of engineers in 2019 with the goal of assisting businesses in transforming their lengthy written content into engaging short videos. The creators of Pictory AI understood that many businesses struggle with creating videos due to limited resources, technical know-how, or budget constraints. To address these challenges, they designed Pictory AI as an intuitive and accessible solution that can be utilized by anyone, regardless of their level of expertise. By harnessing the capabilities of AI, Pictory AI automates and streamlines the video creation process, saving businesses valuable time and money. By using Pictory, users can input their written content, such as blog posts, articles, or other textual materials, into the tool. The advanced AI algorithms embedded within Pictory AI analyze the content and automatically generate a video script and storyboard based on the key points and structure of the input. This automated process significantly reduces the time and effort required to create compelling videos. Pictory served as a related work in our project, playing an important role in our understanding and implementation of our development process.

4.Expected Achievements

Our project aims to create a user friendly and efficient platform for generating high quality content, particularly short videos, without user prior experience. by leveraging the power of multiple AI APIs and combining them together for one complete video.

4.1 Success Criteria

- Generate smart content accurately and in reasonable time.
- Positive user feedback and satisfaction.
- User friendly interface for easy video creation.
- Seamless integration between the AI tools.
- Handle with large user base.

4.2 Unique Features

- Users have the ability to customize the generated content.
- AI driven scrip generation.
- Real time preview and feedback.
- Integration of multiple AI Tools.
- Free accessibility.

5. Engineering Process

In this section we will cover the main steps we identified in our project, as well as the potential challenges that we may encounter during the during the engineering process of the product.

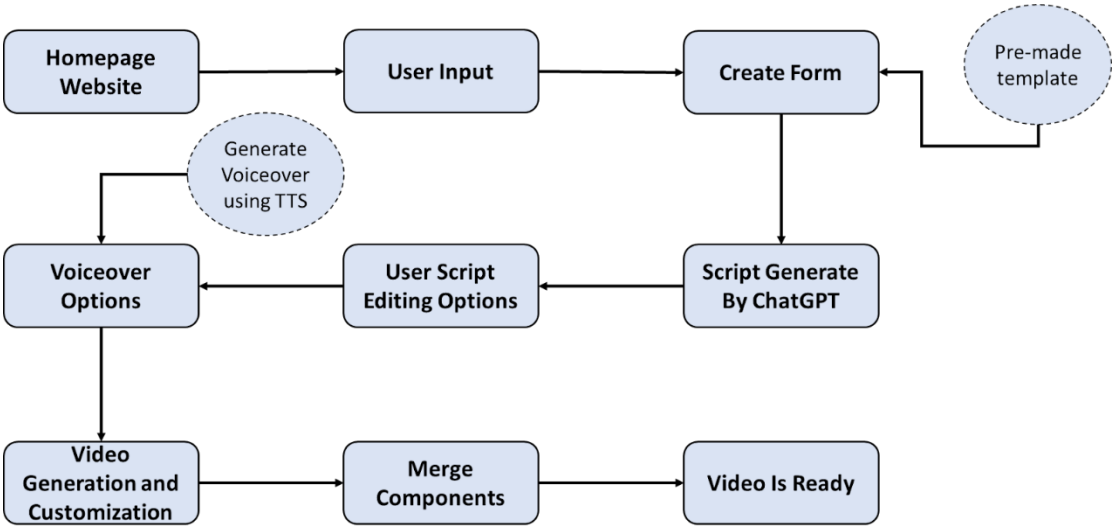

Figure 5: System Architecture

5.1 Process

5.1.1 Website Page

Our website serves as the gateway to our application, we aim to develop a user interface that is intuitive and easy to navigate, even for users without any previous experience or technical expertise. To create a welcoming atmosphere on the website, we will apply a clean and modern design approach with strategically placed elements. that approach helps to achieve a visually appealing interface that avoids overwhelming the user. The design will remain consistent throughout all the video creation steps, ensuring that users can seamlessly progress from one step to the next. In order to optimize the process of developing the website we will use tools such as Bootstrap and React.

5.1.1.1 React

React (also known as React.js or ReactJS) is an open-source, frontend, JavaScript library for building user interfaces or UI components. It is maintained by Facebook and a community of individual developers and companies. React can be used as a base in the development of single-page or mobile applications.
React allows developers to create reusable components and efficiently update them based on data changes. It uses a virtual DOM for fast rendering and follows a declarative approach and enables the creation of dynamic and interactive UIs.

5.1.1.2 Bootstrap

Bootstrap is a free front-end framework for faster and easier web development, Bootstrap includes HTML and CSS based design templates for typography, forms, buttons, tables, navigation, modals, image carousels and many other, as well as optional JavaScript plugins Bootstrap also gives you the ability to easily create responsive designs.

5.1.2 Create New Video Project

The initial step of the video creation process is to gather essential details from the user through a user information form. This form will prompt the user to provide necessary inputs to personalize their video. Some of the key details to be collected in this form include:

- Topic or theme: Users will be asked to specify the subject or theme they want the video to revolve around.
- Keywords or text: Users can provide keywords and text. This will help to provide a script optimally built for their requirements.
- Tone or mood: Users can indicate the desired tone or mood of the video, such as serious, playful, informative, or emotional.
- Text length: Users can provide an approximate length or duration for video.

5.1.3 Integrate User Input with Pre-Made Form.

Once the user provides their personalized details through the user information form, the next step is to integrate that input with a pre-prepared template form that allows us to harness the power of ChatGPT AI language model to generate an engaging script tailored to the user's preferences. The pre-prepared template must be effective and fine-tuned in order to get a high-quality script that leverages ChatGPT capabilities and do so for every user input. Whether the user seeks a formal tone, a conversational style, or specific keyword inclusion, our AI-driven script generation process takes these preferences into account, in order to deliver a script that meets the user's expectations.

5.1.3.1 OpenAI API

To use the OpenAI API effectively its essential to follow a systematic approach that ensures seamless and efficient integration, the following steps outline the Process. Our first step is to create an API key from OpenAI. This key is a unique access code to enable secure communication and authentication with the API.

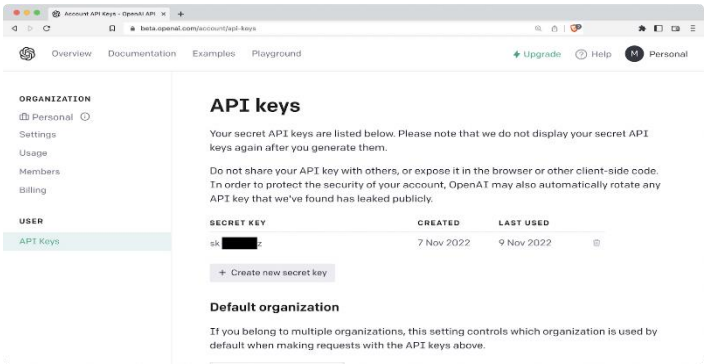

Figure 6: Generating an API Key

After obtaining your API key, the next step is to configure the chat model for the API call. In this step, we need to set up the necessary parameters to interact with the OpenAI API and input any query you want to collect the desire data.

```
import os
import openai

# Load your API key from an environment variable or secret management service
openai.api_key = os.getenv("OPENAI_API_KEY")

chat_completion = openai.ChatCompletion.create(model="gpt-3.5-turbo", messages=[{"role": "user", "content": "Hello world"}])
```

Figure 7: Example of an API Request

By using infinite while loop so that we can chat with the ChatGPT API repeatedly without executing the code again and again. For the final step we need to extract the relevant data from the API response. This data can be further processed, analyzed, and transferred to other AI tools. Also, we can use a JSON file that includes the necessary information for the ChatGPT API request, such as the model to use, the role and content of each message, and any other relevant parameters for your specific use case.

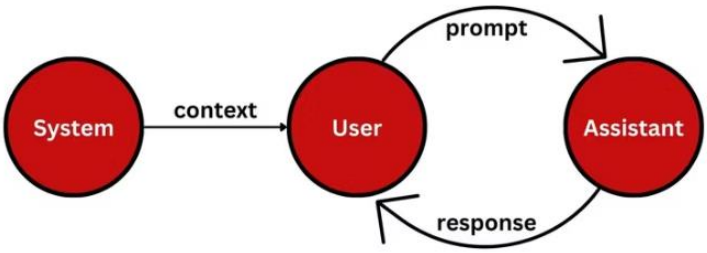

Figure 8: Interaction between ChatGPT and User

5.1.4 Present The Generated Script

After we generate the script by using ChatGPT, we present it to the user and provide them with the option to review, edit and make changes as desired.
The script will be displayed in a clear and readable format, with a user-friendly editing interface that allows the users to refine the script, make adjustments, add specific details, or rewrite sections entirely. By providing this editing capability, we empower users to customize the script to their liking and ensure it aligns perfectly with their requirements.

5.1.5 Generate Voiceover

In this stage we will integrate with AI text-to-speech API to convert the final script into high-quality voiceover for the video, add a human-like voiceover to the visual content, enhancing the overall impact and engagement of the final video.

The users can select from a range of voice options and speech styles, desired tone, gender, language, or accent preference. Whether they prefer a professional and formal voice or a friendly and conversational tone, our application will offer them to choose their specific preference. Being able to hear the voiceover reading the video script helps the user choose the exact voice they want for their project. There are many TTS technologies that offer APIs we can integrate with, like PlayHT and Eden AI, each API has different capabilities, characteristics, and prices. We will have to find the one that suits our requirements, both in terms of performance and cost.

5.1.6 Video Customization

At first, we will use the OpenAI API in order to analyze the user's script and identify several key scenes in it. Then, by understanding the context of each scene, we will generate relevant keywords that encapsulate the desired visual elements. These keywords will be used to search for suitable videos that align with the script and overall narrative, from an extensive database such as Pexels.
Pexels is a popular online platform that provides a vast collection of high-quality videos and images, and its API allows seamless integration of these visual assets into our video creation process. We recognize the importance of customization, and therefore, will provide the users with the option to modify or replace the selected videos for each scene, users can explore the database themselves, searching for videos that better match their vision and preferences. This ensures that the final video reflects the user's creative direction and enables them to curate a visual experience that resonates with their intended message.
Another option that we would like to include in this stage is the ability to create the video using text-to-video AI to generate costume video clip to each scene, however, the accessibility and reliability of such technology remains uncertain at present.

5.1.7 Edite and Merge

By using video editing we can arrange the selected video clips in the desired order, trim, cut, and adjust the duration of each clip to fit the scene's timing and pacing. Finally, the last step is merging the selected videos, along with the voiceover and subtitles to create a cohesive and captivating video that truly brings the user's vision to life.

5.1.8 Video Is Ready

Once the editing is complete, we render the final video file, optimizing it for different formats and resolutions to suit various platforms and devices, the user can then preview the video and ultimately obtain the completed video ready for download or sharing. In case the user is unsatisfied with the generated video he can always go back to a previous step and edit it has he wish.

5.2 Methodology and Development process

The methodology we choose for our project is Agile. This methodology highlights iterative development, collaboration, and adaptability.
By using the Agile methodology, our main goal is to encourage teamwork and adaptability. Agile allows us to work in small, manageable Sprints and easily adjust to changes in the project. The development process:

- API Research and selection: we need to evaluate different AI APIs that suit with our project requirements and select the most suitable APIs based on performance, features and how smooth the integration process.
- Creating premade form: we will make a form that contains specific questions for understanding the user intent efficiently, for example: Topic, Video time, Voice, Key words and more.
- Integrate and Preprocessing: integrating the user answers with our premade form and sending it to ChatGPT to improve and make the form more efficient for sending it to other AI tools.
- User interface Design and development: we need to design and develop an intuitive and user-friendly interface that anyone can use it, for example: content preview, customization options, big buttons etc.
- Testing and quality: Functional testing, usability testing and performance testing, we need to resolve any issues that arise during the testing is part of our development.
- Feedback: we will gather user feedbacks to identify areas for improvement.

In the end of each Sprint period, we will evaluate if there are any necessary changes needed to be made before we continue to the next Sprint.

5.3 Constraints and challenges

Creating this project comes with its fair share of challenges that we need to overcome. These obstacles will test our problem-solving skills and require careful consideration to ensure the success of our project. By addressing these challenges, we can ensure a smoother, more successful, and more efficient development process.

5.3.1 Integration with AI Tools and APIs

Integrating with AI tools and APIs, such as ChatGPT, requires careful implementation and management. Each AI tool has his specific requirements or dependencies that we need to consider. We may face challenges related to compatibility, versioning, and working within the limitations of the AI tools and APIs we choose to use. Therefore, staying updated with the latest versions and ensuring seamless communication and data exchange with these tools is essential to unlock the full potential of AI and optimize the integration.

5.3.2 User Input Variability

We may face the issue of handling the variability of user input. The user may provide input in different formats and styles, which can make it difficult to accurately capture and understand their intent. Each user can have their own preferences and requirements, and they may express them in different ways. This diversity in user input adds complexity to our project. To overcome this challenge, we will provide clear guidance and validate the user input. Moreover, To simplify and make more accurate input we will use check boxes and dropdowns menus in the appropriate places.

5.3.3 Real Time performance

Generating content in real-time and delivering it within acceptable response times can be very challenging, especially when handling a large number of users simultaneously. When handling a significant volume of user requests simultaneously, the system may experience increased computational load, resulting in delays or slower response times. This can impact the user experience. To address this challenge, we need to optimize our system structure, using efficient algorithms, and splitting the workload to multiple servers for making a smooth user experience, even when there are many users accessing the system at the same time.

5.3.4 Quality Control and Filtering

Ensuring that the generated content meets quality standards and filters out inappropriate or undesirable content can be a challenge. It requires implementing strong quality control measures, content filtering algorithms, and moderation processes. By effectively filtering and monitoring the generated content, we can guarantee that it meets high-quality standards and aligns with our guidelines.

5.3.5 Accuracy and Quality of Generated Content

Ensuring that the AI mode produces accurate and high-quality content that meets the user’s expectations can be challenging. As we mentioned before AI models may sometimes generate content that is irrelevant , nonsensical or contains errors. In order to address this challenge, our project use a pre-built form that enables users to provide detailed input and guidelines. By capturing specific requirements and preferences from the users, we can optimize the content generation process.

6.Product

6.1 Requirements

Functional Requirements:

1.	Allow users to input their content preferences and requirements such as: Topic, video time, style, and specific requirements into a form
2.	Our system supports an input script from the user
3.	The system provides options for the user to customize and edit the generated content.
4.	Content filtering to prevent the generation of inappropriate content.
5.	The system integrates with AI APIs tools.
6.	Sharing the generated content with popular social media platforms for direct publishing such as: Facebook, Instagram etc.
7.	The system generate voiceover

Non-Functional Requirements:

1.	processing user input and generating content within acceptable response times
2.	The system designed to handle a large number of concurrent users without compromising performance.
3.	The system measures to protect user data and ensure secure communication with AI APIs tools.
4.	The system needs to be reliable, with minimal downtime.
5.	The system needs to be compatible with different devices and browsers,
6.	The system generates accurate and high-quality content that aligns with user preferences and requirements.

6.2 Diagrams

Diagrams are an important tool for understanding how the system works. These diagrams help us visualize and clarify the system behavior, making it easier to design and explain how things should work. They are like maps that guide us in understanding and communicating the different parts and interactions of the system.

6.2.1 Use Case

The following Use Case shows the user integration with our system.

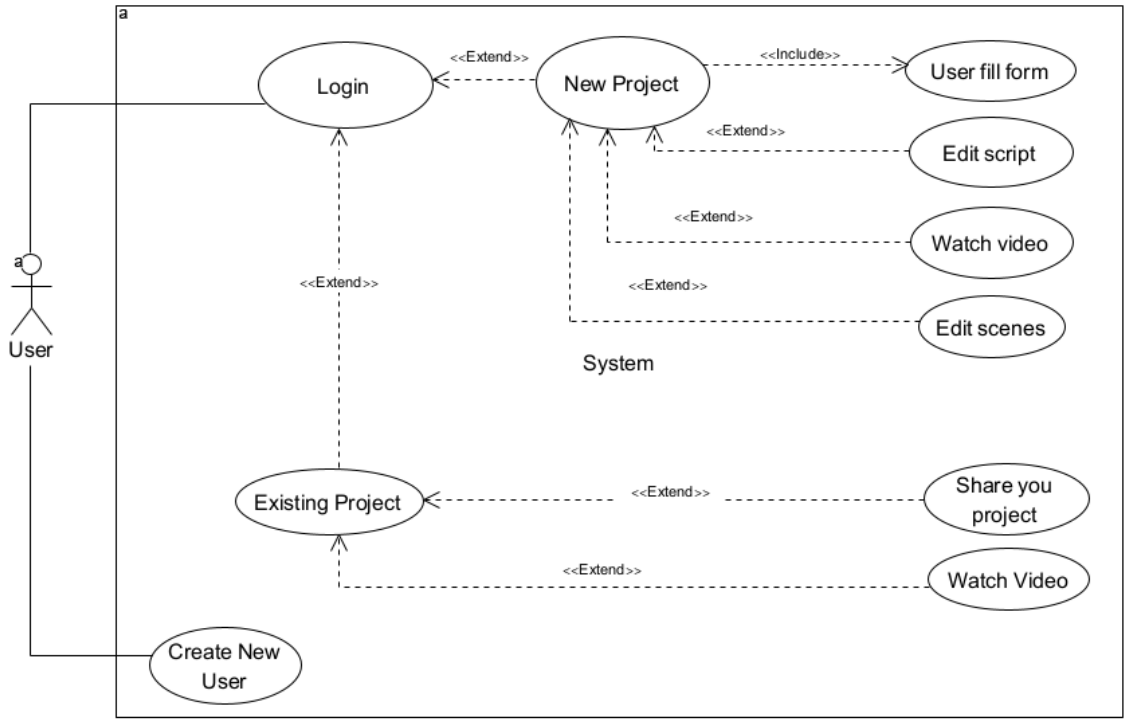

Figure 9: Use Case Diagram

6.2.2 Activity Diagram

The following diagram describe the entire process starting with the user enter to the website until the video is ready to watch.

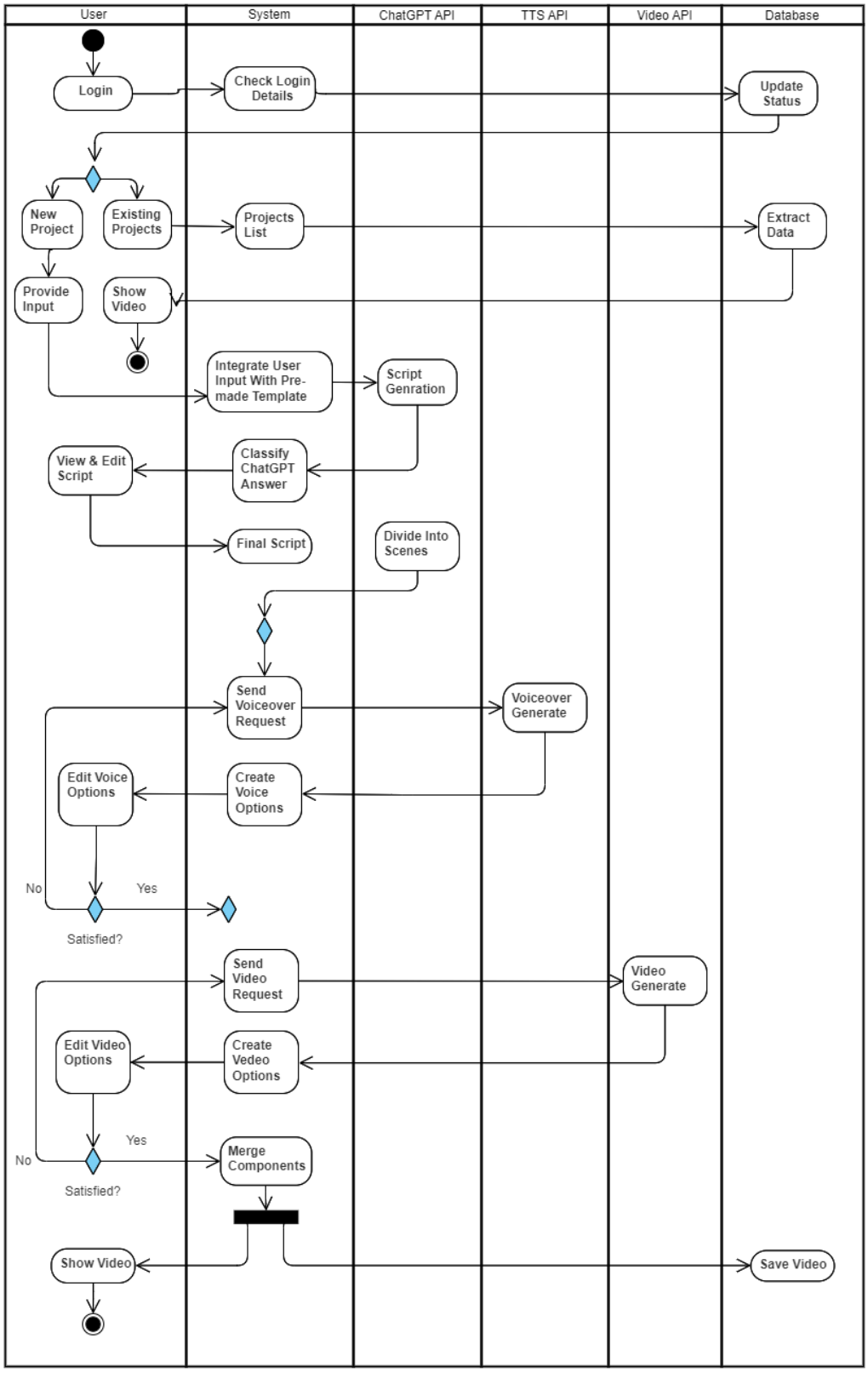

Figure 10: Activity Diagram

6.3 Development

- In our project for the Frontend development, we will use HTML, CSS, and JavaScript to create and design the user interface. Additionally, we will use JavaScript library such as React.js, Angular for simplify the process.
- On the Backend side, we will use Python or java for handling server side and interact with the database.
- Build Git repository for Version control. [Link]
- Use MongoDB as a database for saving data such as videos, scripts, and user details.

6.4 User Interface

6.4.1 User Login Screen

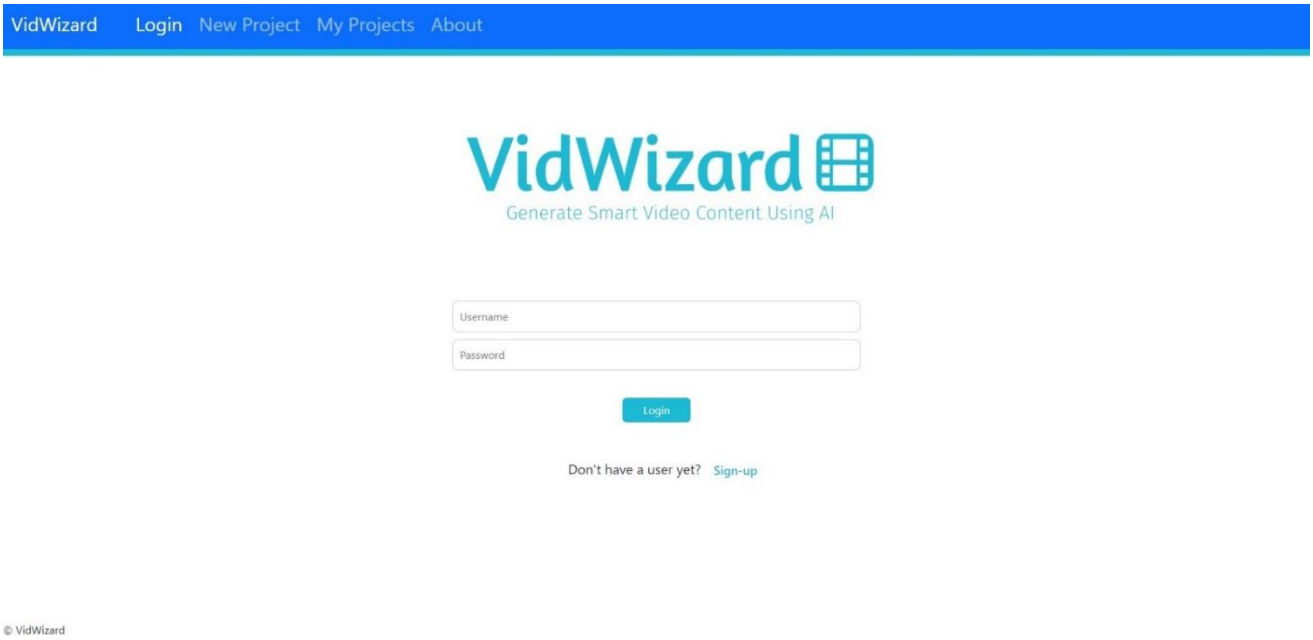

Figure 11: User Login Screen

6.4.2 User Input Screen

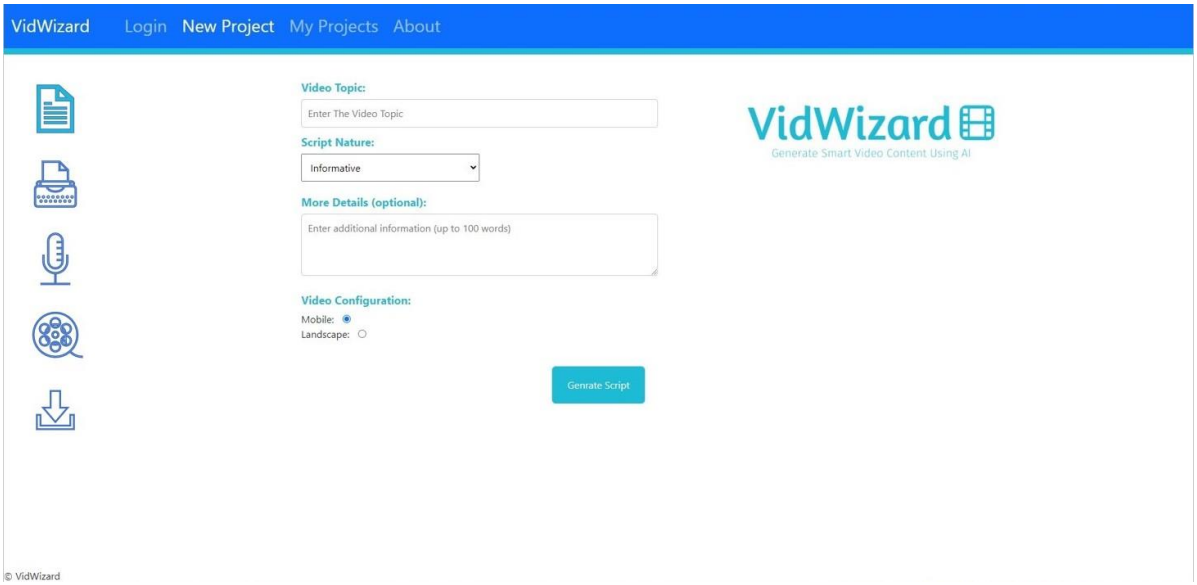

Figure 12: User New Project Input Screen

7. Verification and Evaluation

7.1 Testing Plan

We have divided our testing plan into two parts: the Frontend testing and the Backend testing.

7.1.1 Frontend Testing

1. We are planning to use a testing framework JestJS or Selenium, with those frameworks we can perform functional testing by writing test cases to ensure the user interface is working correctly for example input forms, buttons, menus and more.

2. We will test our system across different browsers, devices, and screen sizes to ensure consistent and optimized user experience. This can help us detect issues that may arise in specific environments.

Figure 13: Example of JestJS

7.1.2 Backend Testing

In the Backend part we first test our Database to verify that the data is correctly
stored and updated in the SQL, by testing different operation with the help of tools such as SQLUnit or DbUnit. Secondly, we will test the integration with the AI APIs to ensure proper communication and accurate results within a reasonable response time. tools that we can use for the APIs testing: Postman, REST assured library.

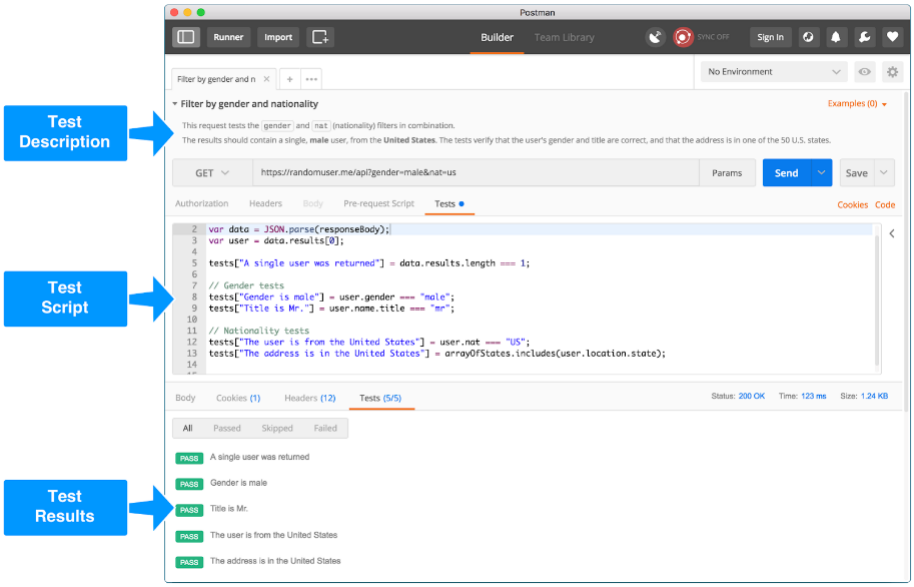

Figure 14: Example of Postman

Furthermore, we would like to test our UI on group of people and ask for their feedback about the use of our interface by making surveys.

7.2 Evaluation

In order to evaluate our platform, we will let users fill out surveys about the overall experience working with our platform and their satisfaction level from the final result. In addition, we will provide the opportunity to additional free text feedback regarding their user experience or any faulty features within our platform.

8.References

[1] Copeland, B. (2023, June 21). Artificial intelligence (AI) | Definition, Examples, Types, Applications, Companies, & Facts. Encyclopedia Britannica. https://www.britannica.com/technology/artificial-intelligence (Page 1)

[2] Wikipedia contributors. (2023). Machine learning. Wikipedia. Machine learning - Wikipedia (page 1)

[3] Wikipedia contributors. (2023a). API. Wikipedia. API - Wikipedia

[4] Wikipedia contributors. (2023a). Text-to-video model. Wikipedia. Text-to-video model - Wikipedia

[5] The state of AI in 2021. (2021, December 8). McKinsey & Company. Global survey: The state of AI in 2021 | McKinsey

[6] Gregory, M. (2020, January 15). What Does Fairness in AI Mean? Forbes. What Does Fairness in AI Mean? (forbes.com)

[7] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on Fairness, Accountability and Transparency *(pp. 77-91). PMLR. buolamwini18a.pdf (mlr.press)*

[8] Jeffrey Dastin, (2018, October 11). Amazon scraps secret AI recruiting tool that showed bias against women. Amazon scraps secret AI recruiting tool that showed bias against women | Reuters

[9] Telford, T. (2019, November 11). Apple Card algorithm sparks gender bias allegations against Goldman Sachs. *Washington Post*. Apple Card algorithm sparks gender bias inquiry - The Washington Post

[10] AI-Generated Content (AIGC): A Survey (Jiayang Wu, Wensheng Gan, Zefeng Chen, Shicheng Wan, and Hong Lin) Mar 2023. [2304.06632] AI-Generated Content (AIGC): A Survey (arxiv.org)

[11] AI-Generated Content, Creative Freelance Work and Hospitality and Tourism Marketing (Haaga-Helia University of Applied Sciences, Pajuniityntie 11, 00320 Helsinki, Finland)