

# POLITECNICO DI TORINO

MASTER's Degree in Engineering & Management



MASTER's Degree Thesis

## The Use of Artificial Intelligence Techniques for the Creation of Soundtracks from Videos

Supervisors

Prof. Luca MASTROGIACOMO

Prof. Elisabetta RAGUSEO

Candidate

Enis Yigit TURGAY

July 2024

## **Abstract**

Artificial intelligence is increasingly gaining popularity and finding its role in our daily activities. Although businesses have adopted AI, it is also becoming more powerful in terms of creativity through applications like text-to-speech, text-to-image, and video-to-text. This thesis primarily aims to investigate the use of artificial intelligence in generating soundtracks for videos. The first part of the thesis conducts a comprehensive analysis of the influence of AI on the music industry. It focuses on developments made in machine learning, specifically in video analysis and music generation. Additionally, it includes a case study of three important industry players: Meta, Google, and Open AI. The main focus of the thesis revolves around formulating a business model for a theoretical startup that uses artificial intelligence to analyze video content and produce soundtracks that align with the emotions, themes, and actions shown in the video. The business model consists of defining the inputs and outputs of the system, conducting market research, establishing objectives for potential customers and the market, specifying the constraints of AI, including legal and ethical considerations, and defining the customer and technical prerequisites. Soundtrackr is a start-up that aims to facilitate the process of creating soundtracks by automating manual tasks and providing assistance to users throughout the production stages. Its goal is to save time, enhance quality, and reduce costs compared to conventional soundtrack production methods.







# Acknowledgements

I want to express my heartfelt gratitude to all those who have supported me since day one, and to myself for overcoming every obstacle with unwavering determination and resilience.



# Table of Contents

<b>List of Figures</b>	VII
<b>Acronyms</b>	X
<b>1 Introduction</b>	1
1.1 AI & Music Market Overview . . . . .	1
1.2 Problem Definition . . . . .	2
1.3 Proposed Solution: Soundtrackr . . . . .	3
1.4 Thesis Structure . . . . .	4
<b>2 Bibliographic Study</b>	5
2.1 From Artificial Intelligence to Generative AI . . . . .	5
2.2 Machine Learning Architectures Overview . . . . .	6
2.3 Machine Learning Applications in Video & Music . . . . .	7
2.3.1 Video Applications . . . . .	8
2.3.2 Music Applications . . . . .	8
2.4 Foundation Models and Some Generative AI Companies . . . . .	9
2.4.1 Meta AudioCraft . . . . .	9
2.4.2 Google MusicLM . . . . .	11
2.4.3 Open AI . . . . .	11
<b>3 Business Model of Soundtrackr</b>	13
3.1 Executive Summary . . . . .	13
3.1.1 Our Essence . . . . .	13
3.1.2 What Do Our Customers Want . . . . .	14
3.1.3 Business Area Attractiveness . . . . .	14
3.1.4 Our Objectives . . . . .	14
3.1.5 Key to Success . . . . .	14
3.2 Company Description . . . . .	15
3.2.1 Company Mission Statement . . . . .	15
3.2.2 Service Description and Unique Value proposition . . . . .	15

3.2.3	Customer Segments . . . . .	16
3.2.4	Need of Customers . . . . .	16
3.2.5	The Price of the Service . . . . .	19
3.2.6	Boundaries of Sountrackr's activities . . . . .	19
3.3	System Plan . . . . .	20
3.3.1	Use Case Diagram . . . . .	20
3.3.2	Activity Diagram . . . . .	20
3.3.3	Sequence Diagram . . . . .	20
3.4	Strategic Plan . . . . .	24
3.4.1	External Analysis . . . . .	24
3.4.2	Internal Analysis . . . . .	32
3.5	Development Life Cycle . . . . .	36
3.5.1	Introduction . . . . .	36
3.5.2	Customer Requirements . . . . .	36
3.5.3	Competitive Benchmarking . . . . .	39
3.5.4	Technical Characteristics . . . . .	45
3.5.5	Relationship Matrix . . . . .	47
3.5.6	House of Quality . . . . .	48
3.5.7	Lyman's Normalization . . . . .	50
3.5.8	IDEF0 . . . . .	50
3.6	Organizational Plan . . . . .	53
3.7	Project Plan . . . . .	54
<b>4</b>	<b>Conclusion</b> . . . . .	<b>57</b>
4.1	AI Overview . . . . .	57
4.2	Business Model Conclusion . . . . .	58
<b>A</b>	<b>Survey Questions</b> . . . . .	<b>59</b>
	<b>Bibliography</b> . . . . .	<b>67</b>

# List of Figures

1.1	Overview of AI companies in the audio sector. . . . .	1
1.2	Process of the Soundtrackr. . . . .	4
2.1	Machine Learning vs Deep Learning [8]. . . . .	6
2.2	Machine Learning vs Deep Learning [12]. . . . .	7
2.3	MusicGen and AudioGen Overview [17]. . . . .	10
2.4	MusicGen Overview [17]. . . . .	10
2.5	Painting Caption Conditioning by MusicLM [19] . . . . .	11
3.1	Soundtrackr Logo . . . . .	13
3.2	Soundtrackr Representation . . . . .	15
3.3	Use Case Diagram . . . . .	21
3.4	Activity Diagram . . . . .	22
3.5	Sequence Diagram . . . . .	23
3.6	Business Model Canvas of Soundtrackr . . . . .	32
3.7	VRIO Analysis . . . . .	33
3.8	Gap Analysis . . . . .	34
3.9	Risk Analysis . . . . .	35
3.10	Customer Requirement Ratings and Importance Measures . . . . .	39
3.11	Customer Requirements vs. AVG of Absolute Importance . . . . .	39
3.12	AIVA CR . . . . .	40
3.13	Studio by MatchTune CR . . . . .	41
3.14	Dynascore CR . . . . .	41
3.15	Soundful CR . . . . .	42
3.16	User - Friendly vs Unique . . . . .	42
3.17	Price vs Product Customization . . . . .	43
3.18	Competitor Benchmarking . . . . .	44
3.19	Relationship Matrix . . . . .	48
3.20	House of Quality . . . . .	49
3.21	Lyman's Normalization . . . . .	50
3.22	IDEF0: General View . . . . .	51

3.23 IDEF0: Soundtrackr Overview . . . . .	51
3.24 IDEF0: Detailed View . . . . .	52
3.25 Organizational Chart . . . . .	53
3.26 Roadmap of Soundtrackr . . . . .	55
3.27 Gantt Chart of Soundtrackr . . . . .	56



# Acronyms

**AI**

artificial intelligence

**ML**

Machine Learning

**CNN**

Convolutional Neural Networks

**RNN**

Recurrent Neural Networks

**LSTM**

Long Short-Term Memory

**GAN**

Generative Adversarial Networks

**DAW**

Digital audio workstation



# Chapter 1

## Introduction

### 1.1 AI & Music Market Overview

The intersection of artificial intelligence and the music industry has led to a new era of innovation and transformation. With advancements in machine learning, natural language processing, and audio synthesis, AI technologies are revolutionizing every stage of the music lifecycle. From the creation of original compositions to personalized music recommendations for listeners, AI is reshaping how music is composed, curated, distributed, and consumed. Leading companies and startups are leveraging AI to develop groundbreaking tools and platforms that cater to the evolving needs of musicians, producers, and music enthusiasts alike. The market is continuing to grow as a result of rising demand for content and AI-driven solutions, and it appears that artificial intelligence will have a significant impact on how quickly people consume music in the future. According to the report by

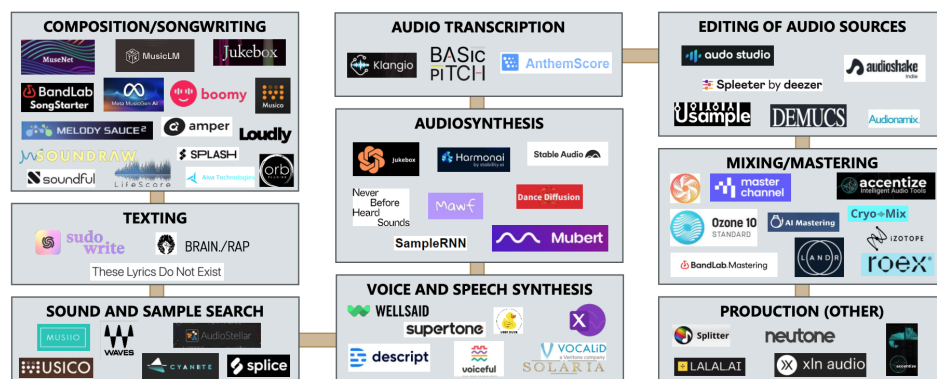


Figure 1.1: Overview of AI companies in the audio sector.

[1]

market.us [2], called "Generative AI in Music Market," between 2023 and 2032, the generative AI in the music market is expected to register a CAGR of 28.6% due to the increasing adoption of smartphones and technological advancements, but there are real concerns regarding the output quality and privacy. Especially on the legal side of generative AI, there are significant questions that need to be answered. Copyright protection is traditionally granted to works created by humans, not machines. This presents a challenge for AI-generated works where the "author" is an AI system. It notes that for a work to be copyrightable, there typically needs to be a significant human creative contribution. The challenge with AI-generated music is determining if the human contribution (e.g., setting parameters or selecting outputs) is substantial enough to warrant copyright ownership. These factors show that there is a need for legal frameworks to adapt to acknowledge and regulate AI's role in creative processes, and according to the article, there should be a collaborative approach involving legal experts, engineers, musicians, and policymakers to explore these issues comprehensively. The goal is to ensure that the development and application of AI in music benefit all stakeholders and respect the legal and ethical norms of society (Sturm, B.L.T et al., 2019) [3]. On the other side, the quality of the output prevents a lot of users from experiencing AI in their music productions, especially in the music generation field. In other fields, such as where AI is just an assistant throughout the music production process, the usage is frequent. According to a recent survey of 1500 people, 11.49% of music producers have found AI music generators useful in their production process, while 22.27% of music producers have never used any AI music tool before. In the meantime, more than 20% of surveyors have concerns regarding the creativity, originality, copyright, and ownership of AI-generated music [4]. Currently, even though there are a lot of doubts about music generation using AI, other AI tools demonstrate remarkable efficiency, creativity, and practicality for music producers. For the music generation itself, with the constant development of legal implications and AI itself, more people will start to use these tools to produce music and get assistance in their production workflow.

## 1.2 Problem Definition

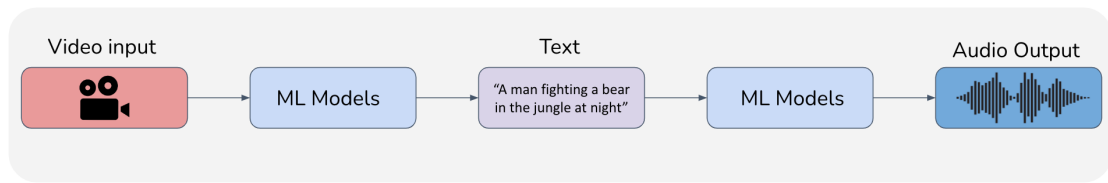
In today's world, technology has changed how content such as video and music is created. Anyone interested in creating content can do so with unlimited resources and powerful electronic equipment. While enjoying the technological advancements of now, the reach of social media is also increasing every day, and in a world where likes and shares are not just numbers, creators feel this pressure to keep pumping out content, even if it means sacrificing quality and originality. The same dilemma happens in all media sectors, including video-on-demand over-the-top streaming

services, causing an abundance of movies to watch with generic soundtracks and movie plots. But even with unlimited resources, several problems that are currently present in the media production industry can be addressed:

- **High costs:** Producing custom soundtracks for media such as films, games, and advertisements can be very expensive.
- **Time & efforts constraints:** Traditional video analysis & editing and traditional soundtrack production can be time-consuming and tiring, from the start to the end.
- **Lack of personalization:** In many scenarios, especially in lower-budget productions, generic stock music is used instead of custom compositions. This can lead to a lack of personalization that might not perfectly fit the mood or themes of the video.
- **Access for independent creators:** Smaller production teams and independent creators often lack the resources to commission original soundtracks.
- **Complicated UI:** When people want to create professional content, there is always a learning curve for the professional software that they need to use.

### 1.3 Proposed Solution: Soundtrackr

The proposed solution is to develop software that can analyze videos using AI and generate soundtracks and the software is called **Soundtrackr**. The purpose of Sountrackr towards its customers is to provide them with a quick, easy, and affordable way to assist them in creating soundtracks for their videos. The application would allow customers to specify their preferences for the soundtrack, such as the genre, mood, tempo, and length. The application would then use AI to provide all the tools necessary to generate a soundtrack that is appropriate for the video and the customer's preferences. It would recommend chord progressions, instruments, arrangements, and automation ideas appropriate for the video. In this way, Sountrackr will help its customers save time and money, give more control over the soundtrack with editing tools and inspiration, and use fewer tools for composing. They will also have the possibility of exporting their project to their preferred DAW. The sound generation will be achieved by converting the video uploaded to a text description using machine learning models that can be used to analyze videos and convert those analyses to text. Then, from the text description of the video, sound will be generated using machine learning models specific to this kind of operation. Figure 1.2 shows a high level description of the process that the software should follow.



**Figure 1.2:** Process of the Soundtrackr.

## 1.4 Thesis Structure

This thesis will contain an overview of the AI music market, what kinds of AI tools can be used for music generation and video analysis, and then a business model including strategic, organizational, project plan, and development life cycle for the theoretical start-up called Soundtrackr. At last, a conclusion to the work done will be reported.

# Chapter 2

## Bibliographic Study

### 2.1 From Artificial Intelligence to Generative AI

**Artificial intelligence**, refers to the simulation of human intelligence in programs that can be programmed to think and learn like humans. The objective of artificial intelligence is to perform tasks that usually require human intellect. These tasks include problem solving, pattern recognition, and decision making. The learning process can be either rule-based or experience-based.

**Machine learning**, also known as ML, artificial intelligence (AI) is dedicated to developing computer algorithms that learn automatically from experience and data. Without explicit programming, machine learning lets computers make predictions or judgments based on data. It is mostly about creating and using algorithms to support these decisions and projections. These algorithms are designed to perform better over time as they manage increasing amounts of data, making it more exact and effective. [5].

**Deep Learning**, is a subset of machine learning that involves neural networks with many layers. With these multiple layers, it is possible to perform tasks like image recognition, video processing, speech recognition, and natural language processing. It is known for focusing on structuring a learning process where it can recognize patterns and make decisions like humans (LeCun et al., 2015) [6]. It is important to know the distinction between traditional machine learning and deep learning. In traditional machine learning, the human developer guides the machine on what type of feature to look for. In deep learning, the feature extraction process is fully automated.

**Generative AI**, refers to a category of artificial intelligence that utilizes machine learning techniques to generate new data that resembles the training data. It is particularly known for its ability to create content that is indistinguishable from that created by humans, including text, images, music, and other forms of media

(Cevallos et al., 2023) [7]. Generative models attempt to imitate specific patterns and create themselves without being given explicit rules. The AI itself can recognize the best solution that fits the specific problem.

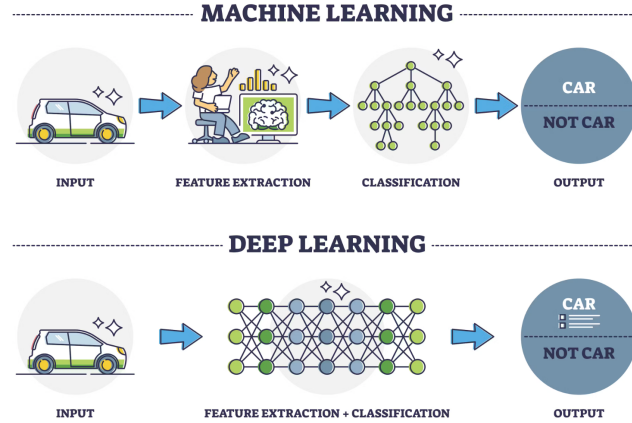


Figure 2.1: Machine Learning vs Deep Learning [8].

## 2.2 Machine Learning Architectures Overview

Machine learning architectures are frameworks that define how algorithms process and learn from data. These architectures determine how data flows through the model, how features are extracted, and how learning is optimized. They are used to solve a range of tasks, from simple classification to complex challenges such as generating sound from video analysis. Below, we explore some of the key architectures that have profound impacts on the field of machine learning.

**Convolutional Neural Networks (CNNs):** Convolutional Neural Networks are a class of deep learning model mainly designed for tasks like object recognition, including classification, detection and segmentation. They excel in processing unstructured data like images, text, audio, and speech [9].

**Recurrent Neural Networks (RNNs):** Recurrent Neural Networks are a type of neural network specifically designed to handle sequence data. They form loops, allowing information to persist (J.Wang et al., 2022) [10]. They take information from their previous inputs so that their output is dependent on the previous computations. RNNs have internal loops that allow them to maintain information in 'memory' over time, making them ideal for tasks where the sequence of data is critical.

**Long Short-Term Memory (LSTM):** Long Short-Term Memory Networks are a specialized type of RNNs designed to address the limitations of traditional



value for the media and entertainment industry [13]. The detailed graphical representations will later be detailed in the Business Plan Chapter. The applications that can be useful for video analysis and sound generation are listed below:

### 2.3.1 Video Applications

- **Object Detection, Recognition & Scene Segmentation:** ML models that identify objects, people, scenes, and actions within videos. Commonly used models for this kind of application are Convolutional Neural Networks (CNNs). CNNs could be used to identify and locate objects within video frames and classify different scenes within a video based on visual features. Also, Recurrent Neural Networks (RNNs) with Long Short-Term Memory Networks (LSTMs) are adept at handling sequential data like data frames, which plays a crucial role to recognize the actions taking place.
- **Content Classification:** Videos can be automatically classified into categories. For this application, CNNs can be used combining Recurrent Neural Networks RNNs or LSTMs to analyze sequences for content categorization.
- **Automated Editing:** Machine learning models can help in editing videos by automatically selecting significant scenes, reducing background noise, and enhancing video quality. Sequence models like LSTMs can identify and assemble significant scenes, while Generative Adversarial Networks (GANs) might be used for creating or modifying content.

### 2.3.2 Music Applications

There are several machine learning models suitable for music production each with their own strengths and weaknesses. The most common approaches are:

- **Automated Composition:** AI-driven tools can generate musical pieces in various styles, supporting composers by suggesting melodies and harmonies. LSTMs and Transformer models are popular for generating music by learning patterns and structures from large datasets of music. Given their ability to process sequences and remember long-term dependencies, LSTMs are a good choice for producing more complex and coherent musical sequences (A.Magar et al., 2023) [14].
- **Sound Mixing and Mastering:** Machine learning is used to automate the mixing and mastering processes, making them more efficient and accessible to independent artists. Deep neural networks are used to automatically adjust levels and balance tracks, applying techniques learned from large datasets of professionally mixed music. Convolutional neural networks (CNNs) could be



used for processing spectral data, while recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, could be useful for dealing with the sequential nature of audio as mentioned previously.

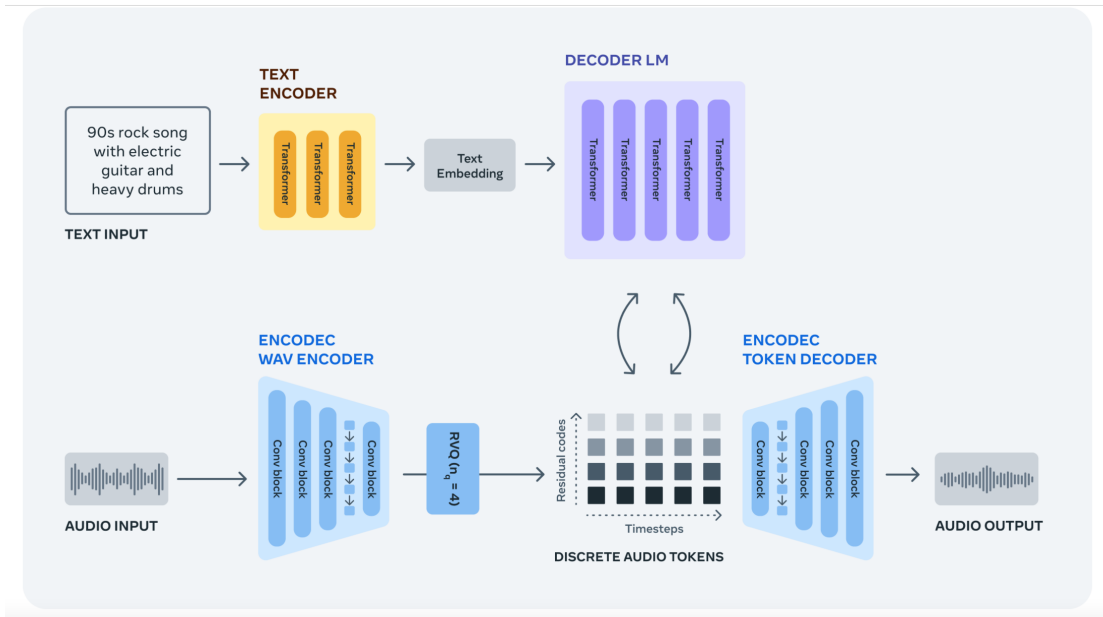
- **Music Classification and Tagging:** Machine learning models classify music by genre, mood, or instrumentation, supporting music discovery and organization. CNNs are employed to classify music tracks based on features extracted from audio (Choi et al., 2016) [15].
- **Dynamic Soundtrack Generation:** In video games and virtual environments, ML models dynamically generate music based on the context of the scene or actions of the user. In interactive environments, Reinforcement Learning can be employed to dynamically generate music that adapts to the context of the scene or user interactions.

## 2.4 Foundation Models and Some Generative AI Companies

Foundation models are large-scale, pre-trained machine learning models developed through deep learning techniques on vast datasets [16]. These models are trained to learn across different types of data, including text, images, video and audio. These models can carry out a wide range of tasks in various areas. Foundation models are often fine-tuned on specific tasks to adapt them to particular domains or applications, for example, for creating music. There are numerous companies creating foundation models which are then used for AI music applications such as MuseNet by OpenAI, MusicGen by Meta and MusicLM by Google.

### 2.4.1 Meta AudioCraft

AudioCraft, created by Meta AI (previously known as Facebook Research), is a unified toolkit aimed at diverse audio generation and processing activities through deep learning. AudioCraft consists of three models: MusicGen, AudioGen, and EnCodec. The purpose of these three models are different from each other. The aim is to have a simple framework that generates realistic audio and music from text-based user input. While AudioGen generates audio from environmental sounds, MusicGen can generate music focused on textual or melodic features. EnCodec is used to compress any kind of audio and reconstruct the original signal with high-fidelity [17].

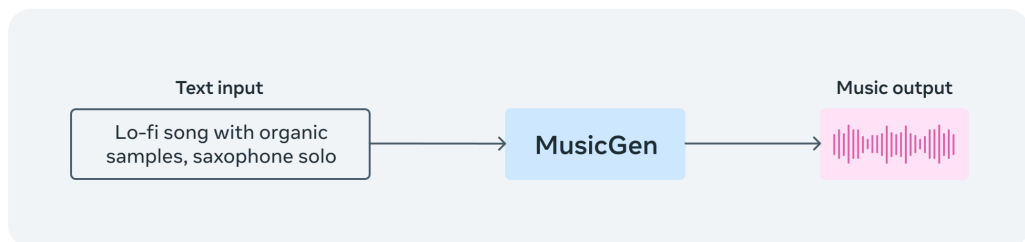


**Figure 2.3:** MusicGen and AudioGen Overview [17].

## Meta MusicGen

MusicGen is an audio generation model that is designed to generate music based on conditions such as textual descriptions or melodic features. This allows users to specify the type of music they want to generate, such as a genre, mood, or even a specific melody line. MusicGen operates using a single transformer model rather than multiple cascading models, simplifying the generation process. Users can influence the output through detailed textual descriptions or by defining specific melodic structures. It can generate high-quality samples, both mono and stereo (J.Copet et al., 2024) [18].

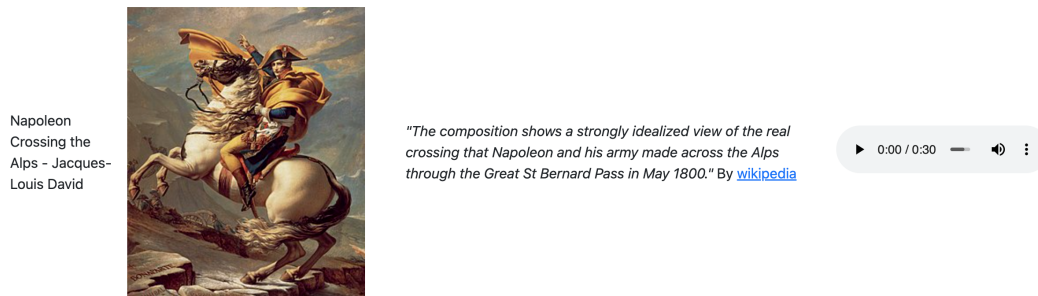
### Text-to-music generation



**Figure 2.4:** MusicGen Overview [17].

## 2.4.2 Google MusicLM

MusicLM is a model developed by Google Research that can generate music based on textual descriptions. It allows users to create unique music pieces simply by describing their desired style, mood, or theme with text. MusicLM can handle text and melody inputs simultaneously, transforming hummed or whistled melodies according to the style described in the text, allowing more creativity.



**Figure 2.5:** Painting Caption Conditioning by MusicLM [19]

It utilizes several neural network architectures to achieve its goals of high-fidelity music generation from text descriptions, such as Transformer models, autoencoder models, MuLan, and Soundstream. The core of MusicLM is built on transformer-based architectures. These are particularly used in the sequence-to-sequence (Seq2Seq) modeling tasks within MusicLM, where transformers handle the generation process by mapping input sequences (text descriptions or audio tokens) to output sequences (music audio tokens). Autoencoders are used for the quantization process that helps in converting raw audio into a discrete representation, which is easier to handle by transformer models in the generation process. Soundstream is a framework that includes neural network components designed to process audio. It serves as an audio codec within MusicLM, and the MuLan model is used for creating joint embeddings of music and text. It employs architectures suitable for handling both audio and text effectively, such as transformers (A. Agostinelli et al., 2023) [20].

## 2.4.3 Open AI

### MuseNet

MuseNet was an AI project developed by OpenAI, designed to generate musical compositions with the ability to combine styles from different composers, ranging from Mozart to the Beatles. MuseNet was capable of creating music with up to 10 different instruments and could generate compositions up to 4 minutes long. It utilizes a deep neural network with 72 different musical styles, showcasing the

versatility of AI in understanding and creating complex compositions [21]. This project highlighted the expanding capabilities of artificial intelligence in creative fields, demonstrating how AI could learn and reproduce the nuances of music theory and style from various genres and eras.

### **Jukebox**

Jukebox is a more advanced neural network model capable of generating music, including rudimentary singing, in various musical styles and genres. Unlike MuseNet, which focused on instrumental compositions, Jukebox can generate music with coherent vocals and lyrics, making it significantly more complex and sophisticated. It works by processing audio data at the waveform level, creating detailed and nuanced musical output. It uses a method called VQ-VAE (Vector Quantized-Variational AutoEncoder) to handle the large variability in audio data efficiently. The model can generate diverse, high-quality songs with coherence lasting several minutes, conditioned on the artist and genre to influence the style, and on lyrics to control the singing (P.Dhariwal et al., 2020) [22]. The model was trained on a large dataset of music from a variety of genres, enabling it to mimic styles ranging from rock and pop to hip-hop and classical. One of the unique features of Jukebox is its ability to continue existing songs or create new compositions that feel stylistically consistent with the work of particular artists or genres. This makes it a powerful tool for exploring new musical possibilities and for understanding the components of different musical styles [23]. The generated music exhibits high fidelity and diversity, demonstrating advancements over prior music generation models that primarily focused on shorter clips or symbolic music representation.

These applications demonstrate the versatility of machine learning in enhancing and automating video and music production, offering significant improvements in both quality and user experience.

## Chapter 3

# Business Model of Soundtrackr

In this chapter, the business plan of a theoretical start-up will be discussed.

### 3.1 Executive Summary

#### 3.1.1 Our Essence

Soundtrackr is a software that enables composers to create groundbreaking songs for their videos. We believe that in a world that is fast-paced and highly competitive, the traditional methods of creating soundtracks can be costly and time-consuming. Soundtrackr can also act as an AI assistant, providing users with a solution to these problems.



Figure 3.1: Soundtrackr Logo

### **3.1.2 What Do Our Customers Want**

Sountrackr is the perfect tool for music composers and content creators who want to create high-quality soundtracks quickly and easily. Here are some additional benefits of using Sountrackr:

- **Cost-effectiveness:** Sountrackr is a more cost-effective way to create soundtracks than traditional methods.
- **Simple to use:** Sountrackr is simple to use, even for those who do not have a background in music.
- **Flexibility:** Sountrackr offers flexibility in creating soundtracks for a wide range of projects, including videos, games, and advertisements.

### **3.1.3 Business Area Attractiveness**

The demand for video and music content is growing rapidly, and AI technology is becoming increasingly accessible and affordable. The market is competitive, and with the right approach, it is possible to gain market share from the current competitors. There is also a great potential for innovation in the field of AI-generated soundtracks.

### **3.1.4 Our Objectives**

- Generating time and cost-efficient soundtracks;
- Assist music composers;
- Increasing the number of outputs;
- Preventing writer's block;
- Enhance accuracy and creativity;
- Gain a competitive advantage in the AI-generated soundtracks field;
- Being at the forefront of innovation.

### **3.1.5 Key to Success**

Sountrackr will use machine learning models to learn through video analysis and it will generate soundtracks from the analysis it does. Also, it will provide all the tools that a composer needs to create a soundtrack. Sountrackr will give inspiration to creators in just a few minutes, bring creativity, and decrease the costs and time spent to allocate resources by the creator due to its constantly growing database.

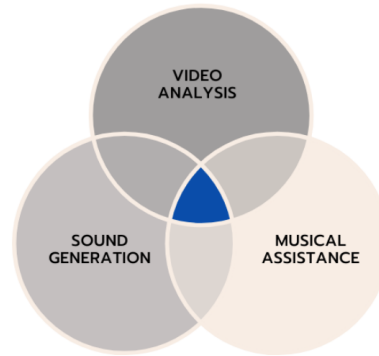


Figure 3.2: Soundtrackr Representation

## 3.2 Company Description

### 3.2.1 Company Mission Statement

The purpose of Sountrackr towards its customers is to provide them with a quick, easy, and affordable way to assist them in creating soundtracks for their videos. The application would allow customers to specify their preferences for the soundtrack, such as the genre, mood, tempo, and length. The application would then use AI to provide all the tools necessary to create a soundtrack that is appropriate for the video and the customer's preferences. It would recommend chord progressions, instruments, arrangements, and automation ideas appropriate for the video. In this way, Sountrackr will help its customers save time and money, give more control over the soundtrack with editing tools and inspiration, and use fewer tools for composing. They will also have the ability to export their project to their preferred DAW.

### 3.2.2 Service Description and Unique Value proposition

Sountrackr is the smartest way to find inspiration and compose songs specific to the videos you upload. The full potential of Soundtrackr can be achieved by creating a user interface that allows users to simply drag and drop or select a video file. While the video is being processed, the user specifies their preferences for the soundtrack, such as genre, mood, tempo, and length. In the background, every detail of the video will be written as text with the video-to-text tools and it will be combined with the users' preferences. Soundtrackr will generate a soundtrack by using text-to-audio tools that are integrated into the application, and it will also give advice on instruments, chord progressions, arrangements, and automation ideas appropriate for the video. The users will then be able to regenerate the

process if they do not like the results. If the results are suitable for the intended audience, then the user will be able to export stems, full songs, or MIDIs. These can help music composers and content creators find new ideas for their soundtracks more quickly and easily.

### **3.2.3 Customer Segments**

It is thought to focus on five customer segments but the most important customers are the music composers and content creators, which help to develop the foundations of the application.

- **Music Composers:** Music composers, who are music enthusiasts who would like to do experiments on the application or people who want to use the application for professional purposes in the most efficient way possible.
- **Content Creators:** Content creators who need to produce a lot of output in a short time period and want to decrease the effort & time spent on music composition.
- **Advertisers:** Advertisers, companies, or individuals who use the app to create custom soundtracks for their marketing videos, aiming to enhance the emotional impact and appeal of their advertisements.
- **Game Developers:** Game developers who utilize the app to automatically generate dynamic soundtracks that adapt to gameplay changes and scenes in their video games, improving the overall gaming experience.
- **Individual Users:** Individual users are casual users interested in creating unique soundtracks for their personal video projects, like family videos, social media content, or just users that would like to experiment with the application, seeking an easy and creative way to enhance their visuals with music.

### **3.2.4 Need of Customers**

This section explores the needs of each customer using the "Pains & Gains" method. This approach identifies the challenges that customers face and how a product can resolve these issues (Pains), as well as the benefits that customers will enjoy by using the service (Gains). Examines the difficulties (pains) that customers currently experience without this solution, with the aim of effectively addressing these with the product. In contrast, the "Gains" part of the analysis highlights the positive impacts and additional advantages customers can expect, enhancing their overall satisfaction and engagement.

#### **Music Composers - Pains**



- Music composers want to finish their projects as soon as possible but struggle to finish them due to writer's block;
- It is time-consuming and expensive to find the right person for the job;
- Lack of inspiration.

### **Music Composer - Gains**

- Music composers are able to find the right tools for their ideas thanks to the countless possibilities generated by the application;
- Music composers are able to finish their projects after getting inspiration from the generated tools;
- The application provides the professional assistance a music composer needs.
- Music composers can save time by automating many of the tasks involved in creating music;

### **Content Creators - Pains**

- Usually, the songs they use have copyright issues.
- Lack of time and resources to create high-quality soundtracks;
- It is time-consuming and expensive to find the right person for the job;
- Lack of inspiration;
- Difficulty finding the right music for their content;
- Difficulty standing out from the competition;

### **Social Media Content Creators - Gains**

- Easily customize music to fit their content;
- Get inspiration from music ideas;
- Easily come up with new ideas by exploring different combinations of instruments and chord progressions;
- Unique and engaging content

### **Advertisers - Pains**

- Lack of time and resources to create high-quality soundtracks;

- Conflict of interests between the company and the hired professional;
- Revisions may increase the project delivery time and the costs of the project;
- Difficulty finding the right music for their commercials.

#### **Advertisers - Gains**

- Companies are able to gain total control of the project, decrease conflicts, and reduce unrealistic costs that may occur during the hiring process;
- Reduced time and upfront costs due to the elimination of the need for musicians and engineers.
- The application provides hundreds of possibilities that allow companies to work on multiple projects at once with little cost.
- Customizable music to fit their commercials.

#### **Game Developers - Pains**

- Usually, the songs they use have copyright issues.
- It is time-consuming and expensive to find the right person for the job;
- Conflict of interests between the company and the hired professional;
- Revisions may increase the project delivery time and the costs of the project;
- Fewer projects to be able to do.

#### **Game Developers - Gains**

- Companies are able to gain total control of the project, decrease conflicts, and reduce unrealistic costs that may occur during the hiring process;
- Reduced time and upfront costs due to the elimination of the need for musicians and engineers.
- The application provides hundreds of possibilities that allow companies to work on multiple projects at once with little cost.

#### **Individual Users - Pains**

- Lack of musical knowledge
- Difficulty creating music that sounds professional.

#### **Individual Users - Gains**

- Easily create high-quality music without any musical skills
- Experiment with different styles and genres of music

### 3.2.5 The Price of the Service

The pricing tiers were developed after analyzing the target customers and competitors. This approach ensures that prices are attractive to customers and competitive in the market, supporting both broad market appeal and sustainability for the start-up.

- **Free Tier:** The company will offer a free tier that allows customers to upload up to 3 videos per month and generate up to 3 soundtracks per month. No monetization and the copyright is owned by Sountrackr.
- **Basic Tier:** The basic tier will be limited to a certain number of soundtracks per video and a certain length of soundtrack. The basic tier will attract a large number of customers who want to dive deeper into the AI-generated soundtracks and want to use the tool just for non-profit and personal use. It will cost 6 Euros per month.
- **Content Creator Tier:** Designed for social media creators, freelancers, businesses, and agencies that need royalty-free background music. It will cost 11 euros per month, allowing customers to upload up to 15 videos per month and generate up to 15 soundtracks per month. This tier provides limited monetization and the copyright is owned by Sountrackr. Be able to export MP3.
- **Pro Tier:** It is for creators who are looking to take their creativity to the next level. The pro tier will cost 30 euros per month and will allow customers to upload an unlimited number of videos and generate an unlimited number of soundtracks. With the pro tier, the customer gets full monetization and copyright. Be able to export high-quality WAV files.
- **Organization Tier:** Licensing & plans for organizations. The company will also offer discounts to businesses and organizations that purchase multiple licenses. The company will also offer discounts to students and educators.

### 3.2.6 Boundaries of Sountrackr's activities

Applications are defined by the technology that is used to develop them. In other words, it will be determined by the algorithms used. This will affect the genre selection and soundtrack length. Moreover, it may not be able to create soundtracks as complex or creative as soundtracks created by humans. Initially, the software will only support videos up to 50 minutes in length. There are also ethical and legal implications, the application should not be used to generate soundtracks that are discriminatory or offensive, and the application may need to obtain permission from the copyright holder of the video before generating a soundtrack for the video.

## **3.3 System Plan**

In the system plan, a general view of how users and the systems are discussed with use case, activity and sequence diagram.

### **3.3.1 Use Case Diagram**

These diagrams help identify the interactions between a system and its external entities. By depicting the various use cases and actors involved, they provide a high-level view of the system's functionality, focusing on what the system does from the perspective of external users. Use case diagrams are particularly useful in capturing the requirements of a system. In the figure 3.3, it identifies roles like User and System Administrator, shows user interactions such as video management and soundtrack editing, highlights system tasks like video processing and administrative controls.

### **3.3.2 Activity Diagram**

These are essentially flowcharts that show the flow of control from one activity to another. They highlight the sequence and conditions for coordinating lower-level behaviors, rather than the behaviors themselves. Activity diagrams are used to model the internal processes of a use case and help in visualizing the operational workflows of a system. In the figure, 3.4, it begins with user login and progresses through actions like video upload and preference setting, features decision points (e.g., validating video) and parallel processes (e.g., AI analysis and soundtrack generation), and concludes with the user downloading the generated soundtrack.

### **3.3.3 Sequence Diagram**

These diagrams detail the interactions between objects and/or parts of the system in the sequence that they occur. They are particularly useful for visualizing how processes operate with one another and in what order. Sequence diagrams describe how and in what order the objects in a system function across a series of activities. In the figure 3.5, it maps out user actions, backend processing, and responses, starting with video upload and ending with soundtrack delivery, and displays the sequence of API calls, database transactions, and data flows necessary for each step.

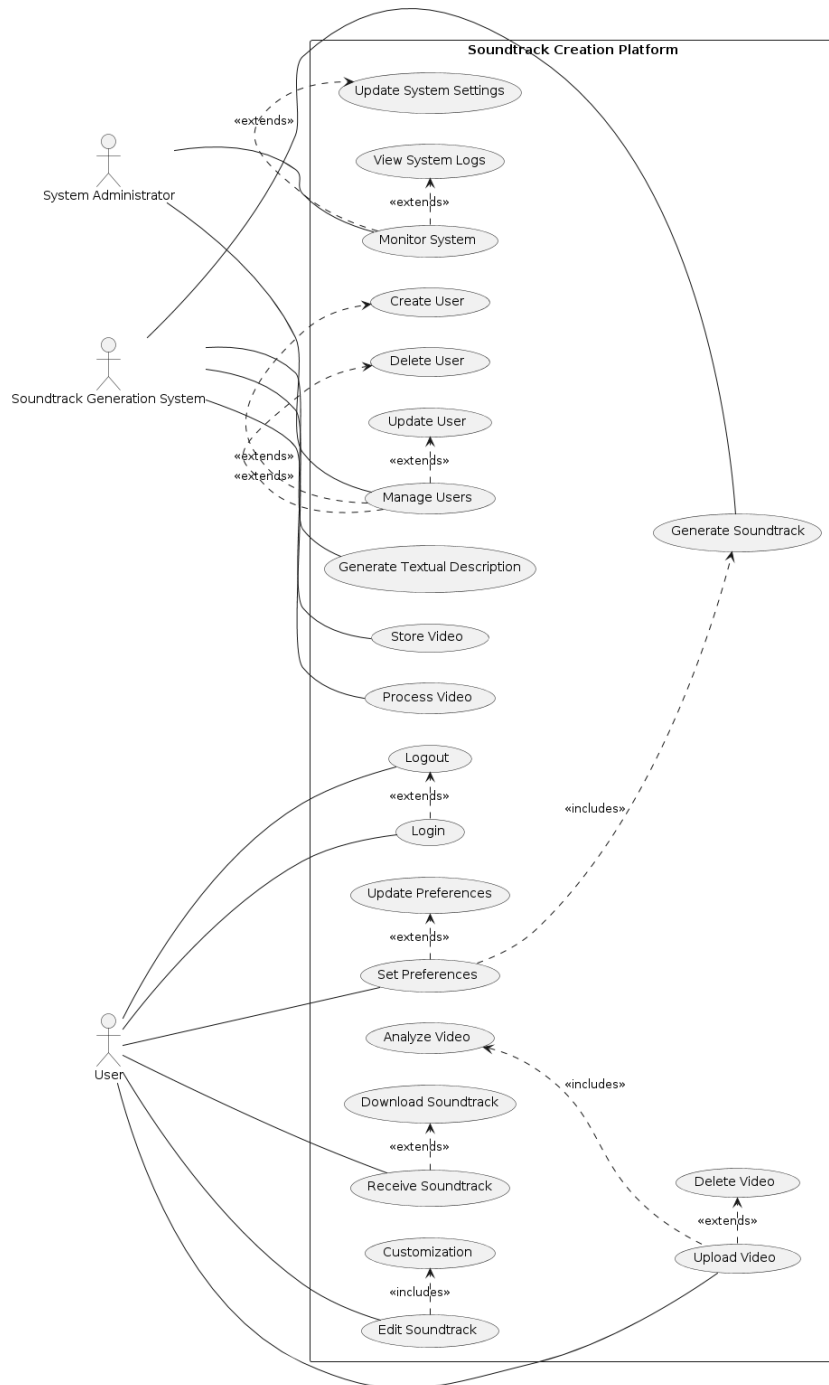


Figure 3.3: Use Case Diagram

### AI-Driven Soundtrack Generation Process

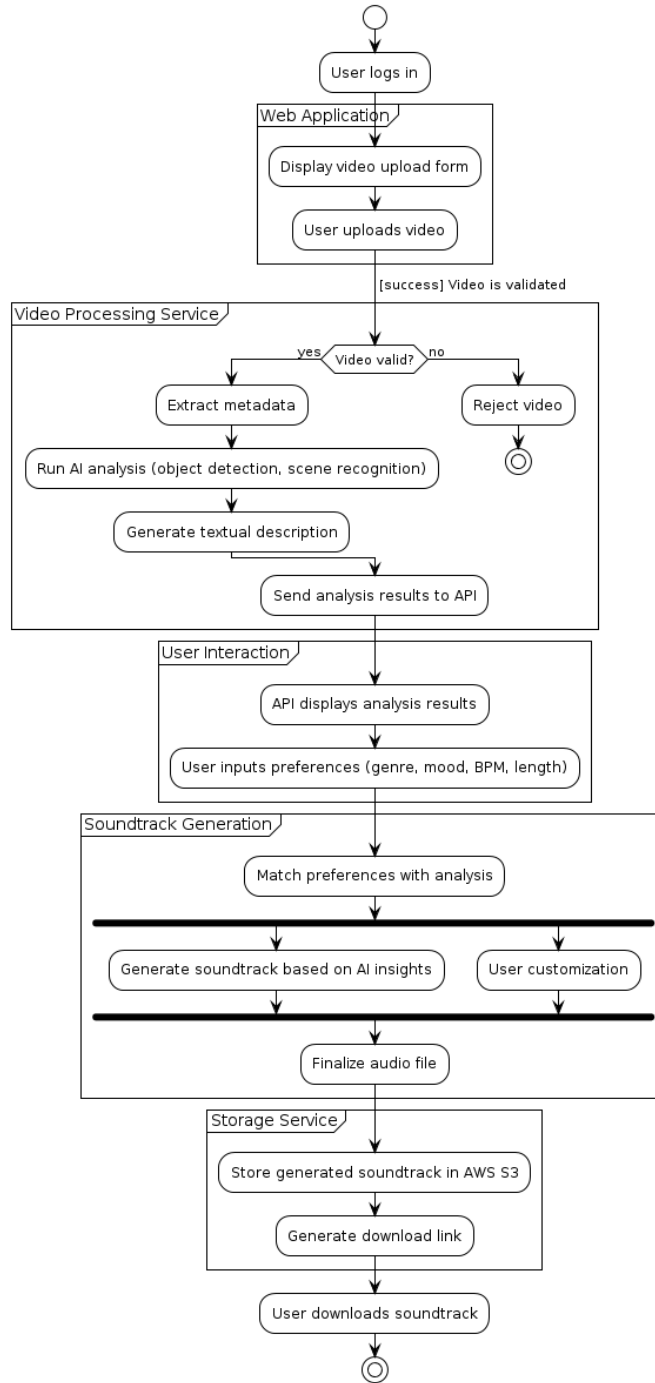


Figure 3.4: Activity Diagram

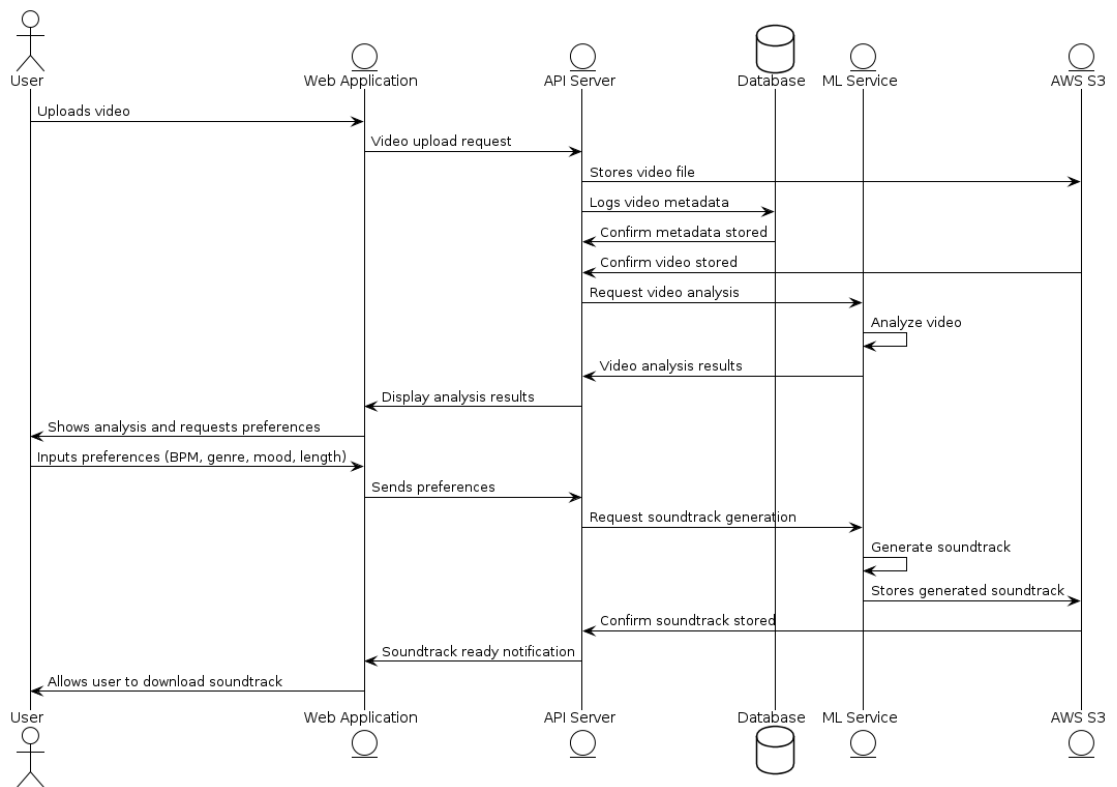


Figure 3.5: Sequence Diagram

## 3.4 Strategic Plan

### 3.4.1 External Analysis

#### PESTEL Analysis

The PESTEL analysis conducted for this startup in the generative AI industry evaluates external factors across Political, Economic, Social, Technological, Environmental, and Legal domains. This comprehensive review identifies challenges and opportunities that could influence strategic decisions and guide the company's growth trajectory.

**Political factors:** The AI market is set to grow and, in parallel, the focus on AI regulations worldwide is increasing every day. Every country has a different approach to this topic, but regulations are fairly standard for generative AI. According to EU AI Act, for generative AI, there are three transparency requirements that need to be complied with:

- Disclosing that the content was generated by AI,
- Designing the model to prevent it from generating illegal content,
- Publishing summaries of copyrighted data used for training.

In the UK, contrary to the draft EU's AI Act, the UK Government does not intend to establish a new AI regulatory body. Instead, it aims to assist existing regulators in applying these principles with their own powers and resources. Proposals for regulation are still under development. In the US, the 'AI Disclosure Act of 2023' mandates that any content produced by artificial intelligence must carry a disclaimer indicating it was generated by AI. Additionally, a significant political aspect involves companies securing permissions from copyright holders before creating soundtracks for videos intended for monetization. [24] [25] [26]

**Economic factors:** The report titled "The economic potential of generative AI: The next productivity frontier" by Mckinsey & Company highlights that the use of generative AI in software engineering within the Media and entertainment sectors could lead to an economic impact ranging from \$580 to \$1200 billion. It also notes an increase in the overall technical potential for expertise in decision-making and collaboration tasks from 24.5% to 58.5% due to generative AI, and a 53% impact on professionals engaged in creative work and arts management. [13].

The "Generative AI in Music Market" report by Market.us states that the market is projected to expand at a CAGR (Compound Annual Growth Rate) of 28.6%, with the software segment capturing a 55% global revenue share in 2022. It is anticipated that the music composition segment will experience the highest CAGR throughout the forecast period. [2].



According to research done on Statista [27], the generative AI market is expected to generate value of approximately US\$45 billion by 2023, with a compound annual growth rate (CAGR) of 24.40%. By 2030, the market is expected to reach a volume of US\$207 billion. Notably, the United States will account for the majority of value generated globally, with US\$16 billion in 2023.

While the potential is there, there are various economic factors that can affect the implementation of this kind of tool.

- **Inflation:** due to inflation costs for labor and benefits, financing, utilities, and even cloud services have increased. Overcoming these costs may damage the products' quality and this will reduce customer satisfaction. Maintenance and updating the product becomes pricey too and it will bring higher fees to the customers. Also, the availability of funding and investment opportunities can affect the company's growth and expansion plans [28] [29] [30].
- **Change in Exchange Rates:** It can impact the company's profitability, especially if it operates globally.

**Social factors:** It is possible that the public may hesitate to accept content generated by AI, especially if they believe that it is not as creative or good as content created by humans. The demand for services related to video content and soundtrack generation may depend on changing consumer preferences and cultural trends. Additionally, it is important for companies to consider ethical and cultural sensitivities when generating soundtracks for videos intended for different regions or communities.

Music composers and producers are already familiar with AI and its capabilities. According to a survey conducted by bedroomproducersblog (BPB), only 17.3% of surveyed producers have negative views about using AI in music production, while almost half (47.9%) have a neutral stance. About 33.6% of surveyed music producers feel that AI music generators that can create entire compositions will only be useful for inspiration. Meanwhile, 30.1% of music producers are planning to try using AI tools soon. However, 15.7% of surveyed producers had already tried AI tools and were disappointed by their quality. AI music generators were ranked as the third most useful tool for surveyed producers, with 11.49% of the vote [4].

The demand for AI in the music industry is on the rise. Meanwhile, the video content market - specifically social media - is experiencing rapid growth. Recent statistics indicate that there will be an estimated 4.89 billion social media users worldwide by 2023, which is a 6.5% increase from the previous year.[31] [32].

**Technological factors:** The development of new AI algorithms that can generate more creative and realistic soundtracks. The business will need to keep up with the latest technological advances in order to stay competitive. The rapid advancements in artificial intelligence (AI) and machine learning technologies are

both opportunities and challenges for the company. On one hand, AI can enable the company to develop more advanced and personalized soundtrack generation algorithms. On the other hand, keeping up with technological advancements and competing with other AI-powered platforms can be a challenge. Additionally, the company needs to consider the technological infrastructure and compatibility requirements of its target market.

**Environmental Factors:** Training an AI model can require significant amounts of computational power, which in turn requires a large amount of energy. This energy is often generated using fossil fuels, leading to an increase in greenhouse gas emissions [33].

**Legal Factors:** The company needs to comply with copyright laws and ensure that it has the necessary licensing and rights to use the music and soundtracks generated or incorporated into its services. Additionally, the company needs to be aware of data protection and privacy laws, especially when it comes to handling user data for personalized soundtrack generation.

### **Porter Five Forces**

The Porter's Five Forces analysis for the generative AI industry provides insights into the competitive environment and market dynamics. It evaluates the threat of new entrants, buyer power, threat of substitutes, competitive rivalry, and supplier power. Each of these forces plays a critical role in determining the strategic direction of a company.

**Threat of New Entrants:** The threat of new entrants is medium. The current competitors are diversified and the brand loyalty of customers can change. Since AI-powered applications of this kind are still in their early stages, regulations are still being proposed. However, building a sophisticated AI system and curating a comprehensive music database for effective soundtrack creation will still require significant investment in research, development, and licensing. Already well-structured companies with well-trained AI models and curated music libraries will have a competitive advantage.

**Buyer Power (Customer Power):** The buyer power is high, as there are many different companies that offer this service. This means that buyers have a lot of choices and can easily switch to a competitor if they are not satisfied with the service, and the switching costs are very low.

**Threat of Substitutes:** The threat of substitutes is high, as there are many different ways to create soundtracks for videos. Video creators can hire a professional musician or sound engineer, or they can use a traditional music library. However, this is time-consuming and costly.

**Competitive Rivalry:** The rivalry among existing firms is high, as there are many different companies that offer AI-powered applications. This means that

companies are constantly competing for customers and trying to differentiate their products.

**Supplier Power:** The bargaining power of suppliers is low, as there are only a few companies that provide the AI algorithms that are needed to develop and launch an AI-powered application. This means that suppliers have a lot of power and can charge high prices for their products. The bargaining power of music suppliers (artists, composers, and music libraries) is moderate. Suppliers hold some power as they provide the raw material (music tracks) necessary for the application's functionality. However, with a large number of music options available, the application developers can negotiate favorable licensing terms. Additionally, the application could explore partnerships with emerging artists or royalty-free music libraries to diversify its supplier base and reduce dependency on a few suppliers.

The company can mitigate the threats posed by these forces by focusing on innovation and differentiation. The company can also develop partnerships with other companies to provide additional features or services to users. By carefully managing these forces, the company can increase its chances of success. Competition in the AI-driven soundtrack creation market is intense. Numerous applications and software platforms offer similar functionalities. Differentiating factors include the quality of generated soundtracks, ease of use, range of available music genres, and integration with video editing tools. Established players with robust AI models and extensive music libraries will have an advantage, making it challenging for new entrants to capture market share. Ongoing innovation and the ability to adapt to evolving user preferences are crucial for maintaining a competitive edge.

### **Key Success Factors**

Key success factors are those variables that management can influence through its decisions and that can significantly affect the overall competitive positions of the company in the industry.

- **User-friendly system:** An easy app for all kinds of customers to get used to the application.
- **Saving money & time:** By not overpaying for sample libraries and reducing the time spent on composing but also easing up the technicalities.
- **Marketing:** The tool should be marketed and promoted effectively to reach its target audience.
- **Continuous improvement:** The tool should be continuously improved to add new features and functionality.
- **Quality of databases:** The database should be updated and monitored very frequently so that the trained AI model has better results.

- **Unique features:** Ability to analyze videos and create soundtracks.

## **Demand & Competitor Analysis**

Understanding market dynamics and competitive positioning is crucial. This analysis identifies the primary customer base and their specific needs, setting the stage for how Soundtrackr meets these demands with innovative solutions. It also explores the competitive environment, highlighting how different players differentiate themselves and the challenges they face in balancing technical complexity with user accessibility. The discussion includes Soundtrackr's strategy to blend uniqueness with user-friendliness, illustrating its potential to redefine industry standards and enhance user engagement.

- **Demand Analysis** The initial target market includes music composers who prioritize creativity over the intricate theoretical and technical aspects of music production. These composers are looking for a digital assistant to streamline their workflow, allowing them to focus on the artistic side of their projects. As the application evolves, it is planned to introduce a sound generator feature to cater to a broader audience, providing a quick, unique, and professional-sounding solution to meet diverse customer needs. Our customers desire a creativity boost, reduced constraints from theoretical knowledge, an easy-to-use interface, continual assistance throughout the production process, and the ability to customize and generate unique ideas efficiently.
- **Competitor Analysis** In the competitive landscape, firms compete by product differentiation, user experience, sophistication of AI algorithms, competitive pricing, and the quality. Most competitors struggle to balance application complexity with user-friendliness, often resulting in powerful tools that are not intuitive for users. Soundtrackr aims to occupy the optimal intersection of uniqueness and user-friendliness. It distinguishes itself by allowing users to upload videos directly onto the platform and automatically generate tailored soundtracks. Furthermore, Soundtrackr's strategy includes targeting a wide customer segment, which enables competitive pricing and offers users extensive possibilities to experiment and 'play' with the application to create the best music for their videos.

## **Competitors**

In this section, the potential competitors are discussed and the potential competitors are: AIVA, Soundful, Studio by Matchtune, Dynascore.

- **AIVA** AIVA is an AI-powered music composition platform that uses deep learning to create original musical pieces. The system analyses the style of

music written by famous composers in history and learns from trials and errors to produce melodies that are pleasing to the human ear. AIVA's AI capabilities enable it to generate musical variations for a given musical theme and then gradually move away from the original theme based on the customer's preferences. Although AIVA has been secretive about its specific algorithm for composing music, the company emphasizes that their goal is not to replace musicians, but to encourage more collaborations between humans and machines. AIVA has been recognized by companies like Vodafone and Nvidia. AIVA's strengths include comprehensive composition from scratch, recognition by well-known companies, and a strong AI team with advanced algorithms for music composition. However, some users may find AIVA's interface and AI-driven composition methods challenging to learn and navigate. Additionally, the high price tag, limited variety of presets, and inhuman-sounding songs may deter some potential customers. AIVA has opportunities to broaden its audience by expanding into other creative industries beyond music, such as film and gaming, which could open new revenue streams. Collaborations with educational institutions to offer AI-driven music composition courses and continuously improving AI models to enhance music composition capabilities and user experience are other potential opportunities. AIVA should also consider DAW compatibility. However, AIVA faces threats such as intense competition from other AI music composition platforms, shifting preferences in music styles and genres, and concerns about data privacy and copyright issues in AI-generated music that could pose legal challenges. AIVA offers several pricing plans for different user needs, categorized mainly for individuals and enterprises. The **Free Plan**, at no cost, is ideal for beginners and requires users to credit AIVA for non-commercial uses, allowing three downloads per month of tracks up to three minutes in either MP3 or MIDI formats. The **Standard Annually Plan**, billed at €11 monthly, targets content creators who wish to monetize their compositions on platforms like YouTube and Instagram, offering 15 downloads per month and tracks up to five minutes long. For more extensive needs, the **Pro Annually Plan** at €33 per month provides full copyright ownership and unrestricted monetization options, with up to 300 downloads per month and the ability to download in all file formats including high-quality WAV files. Enterprises with greater demands can contact AIVA directly for customized solutions tailored to their specific needs and usage. AIVA offers a 30% discount on annual plans and a 15% discount on monthly plans for students and schools. For creative companies that need complete flexibility and intend to heavily commercialize the music created with AIVA, customized enterprise plans are available [34].

- **Soundful** Soundful is a user-friendly website that uses AI to generate unique

and royalty-free music. The platform offers a range of features, including generating, recommending, and editing music. While it has limited music styles and genres, it provides affordable value for podcasters, content creators, and music producers. Opportunities for Soundful include partnerships with video editing software providers and expanding its music library and styles. Threats include competition from established and emerging startups, changes in video content styles, and maintaining high-quality soundtracks that align with video content expectations.

The pricing of Soundful has 6 tiers, from standard to Business Tier 3. Giving multiple possibilities for the users to pick the plan they need. The differences between tiers include the download count, amount of styles and different licensing options [35].

- **Studio by Matchtune**

MatchTune Studio Lite is a web-based audio editing tool that uses artificial intelligence (AI) to assist users in finding and adding music to their videos. It is a popular tool among content creators, YouTubers, and social media influencers. MatchTune Studio Lite is a web-based audio editing tool that uses artificial intelligence (AI) to assist users in finding and adding music to their videos. It is a popular tool among content creators, YouTubers, and social media influencers. MatchTune Studio Lite is renowned for its user-friendly interface, making it accessible even to novices in video editing. It boasts a comprehensive suite of features such as music discovery, synchronization of music with video, and music editing capabilities, all while being cost-effective. This makes it an appealing choice for users looking for good value on a budget. However, the platform has its limitations. Its music library, although growing, does not match the breadth of some competing services. Additionally, the AI-driven music generation can be somewhat repetitive, especially if user preferences are not clearly defined. It's important to note that MatchTune Studio Lite is not a full-fledged video editing software but rather specializes in enhancing videos with music. MatchTune Studio offers different subscription tiers on their pricing page. They have a Free Plan at \$0 which includes watermarked exports and basic features like automatic editing. The Personal Plan is \$9.99 per month, providing enhancements such as no watermarks and social media clearing. For businesses, there's a \$75 monthly plan allowing up to 10 tracks per month and monetization features. They also offer custom Enterprise packages [36].

- **Dynascore** Dynascore is a music composition platform powered by AI that can generate soundtracks for videos, games, and other media. It was developed by Wonder Inventions and released in 2023. Dynascore is unique because it

can create music that is synchronized with video, making it ideal for creating soundtracks for videos and other media. In addition, it offers a variety of features that make it easy to use, even for people with no prior music experience.

This software generates music that syncs with video content, with a user-friendly interface and custom soundtrack capabilities. However, it's expensive and has a steep learning curve, lacking some advanced features. The market for AI-powered music software is growing, with potential expansion into video gaming and virtual reality. Ongoing development can enhance its market presence, but it faces competition from similar tools and changing customer preferences. [37]. Dynascore offers several pricing plans according to their official pricing page:

- **Free Trial:** A 3-day free trial is available, which includes unlimited dynamic tracks and unrestricted usage.
- **iPhone Only:** Priced at \$4.99 per month, this plan includes Dynascore for iPhone with unlimited dynamic tracks and is 100% royalty-free.
- **Individual Plan:** At \$19 per month, this plan provides Dynascore for iPhone, Premiere Pro Extension, and the desktop version. It includes unlimited dynamic tracks and is designed for a single user.
- **Business Plan:** Costing \$39 per user per month, this plan caters to teams of 2-5 users. It includes all the features of the Individual plan along with team collaboration tools.
- **Enterprise Plan:** This plan offers custom pricing for unlimited users and includes all the previously mentioned features, tailored for large teams or companies.

### 3.4.2 Internal Analysis

#### Business Model Canvas

The Business Model Canvas is a strategic tool designed for visualizing, understanding, and shaping a company’s business model. It breaks down key elements like who your partners are, what your business does, and how it earns money into nine easy-to-understand sections. This layout helps companies outline their strategies to better deliver value to their customers, streamline operations, and improve financial performance. Useful for businesses of all sizes, it encourages straightforward planning and quick adjustments to meet changing needs (Osterwalder, A. et al., 2010) [38].

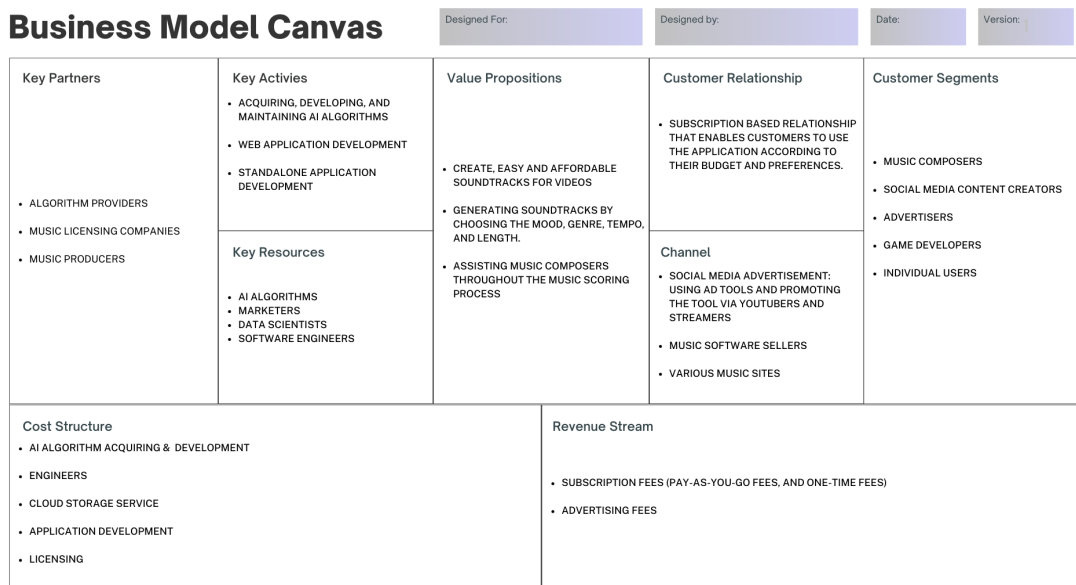


Figure 3.6: Business Model Canvas of Soundtrackr

#### VRIO Analysis

VRIO Analysis is a strategic planning tool for evaluating an organization’s resources and capabilities to discover competitive advantages and potential areas of improvement. Originating from the resource-based view of the firm, the VRIO framework analyzes four dimensions of business resources: Value, Rarity, Imitability, and Organization. Each dimension helps determine whether a resource or capability can sustain competitive advantages.



- **Valuable:** Our product helps users to create music more easily and efficiently that is original and creative and specifically tailored to the results of the video uploaded inside user-friendly software that has exporting options.
- **Rare:** Can analyze videos and create music based on that analysis, which is unique among AI-powered music creation tools.
- **Inimitable:** AI-powered video analysis and music creation tool is inimitable because it is based on proprietary technology that is not available to our competitors.
- **Organized:** Our organizational model is flat. This means there is no middle management between the executives and the frontline employees. This allows us to be responsive when decisions need to be taken. By having this type of organizational structure, we lower operating costs, improve communication, and increase employee motivation and satisfaction.

In the end, it is forecasted a sustainable competitive advantage.

	<b>V</b>	<b>R</b>	<b>I</b>	<b>O</b>
	<b>VALUE</b> Strength that adds value to our customers	<b>RARITY</b> Strength that our competitors don't also possess it	<b>IMITABILITY</b> Strength isn't easily copied by competitors	<b>ORGANIZATION</b> Is our team organized to take advantage of this competitive advantage?
<b>AI - powered video analysis and music creation tool</b>	<b>YES</b>	<b>YES</b>	<b>YES</b>	<b>YES</b>

Figure 3.7: VRIO Analysis

### GAP Analysis

GAP Analysis is a tool used to compare actual business performance with what's ideally expected. It helps identify areas needing improvement to make a company more efficient and meet its goals. In the figure 3.8, there are three focus areas that demonstrate the application's priorities. As the application is currently in its initial stages, there is a 100% gap between its current state and the desired future state. To bridge this gap, several actions need to be taken.

Gap Analysis				
Focus Areas	Desired Future State	Current State	Identified Gap	Action Plan
<i>What are you focused on?</i>	<i>Where would you like to be?</i>	<i>Where are you now?</i>	<i>Difference between desired state and current state</i>	<i>Projects you will undertake to bridge gap</i>
Innovation	To be recognized as one of the most innovative AI-powered video analysis and music creation tool in the industry.	We are not currently known for innovation, the ideas are in place but we need to implement them	100%	<ul style="list-style-type: none"> <li>Add a developer to our core team</li> <li>Try to have more meetings with our target customers in order to understand better how can we create something that they need</li> </ul>
Technology	To develop an algorithm that able to get the information needed from the video analysis, generate a soundtrack and generate musical ideas according to do users' selection and video analysis results	The algorithm has been developed on paper, but still needs to be tested and translated into programming language	100%	<ul style="list-style-type: none"> <li>Add data analysts, database engineers, ML and AI specialist to the team</li> <li>Add music professionals to the team</li> <li>Invest in cloud services</li> <li>Find the best algorithm for the video analysis form the market</li> </ul>
MVP	Built an MVP that works and that serves as a funnel to capture new customers that will be the first users of the final platform	Planned to design a mockup that will allow us to better explain to UX & UI engineers and developers how we want our platform to be seen.	100%	<ul style="list-style-type: none"> <li>Develop the software</li> <li>Test the MVP with real customers and collect their opinions</li> </ul>

Figure 3.8: Gap Analysis

### Risk Analysis

Risk Analysis involves identifying and assessing potential risks to minimize their impact on a company. This process helps businesses prepare for unexpected challenges, ensuring they continue operating smoothly.

As it can be seen from the figure 3.9, the risks that Soundtrackr may face vary from financial risks to risks arising from the internal coordination of the team. These could impact differently depending on their probability of occurrence and the extent of the damage they could cause. Avoiding or mitigating these risks is

therefore of primary importance for Soundtrackr.

There are numerous actions or factors that can contribute to mitigating the risk. Among these, it can be thought of the creation of a heterogeneous and close-knit team (with the recruitment of qualified personnel) as a key element. Having a competent team in various sectors allows us to have a broader understanding of market dynamics and consequently more accurate forecasts; in this way the team can be able to lower the impacts of these risks.

	Risk	Possible effects	Occurrence	Severity
<b>Financial</b>	Insufficient funding to develop and launch the product Unforeseen financial expenses, such as legal fees or patent infringement lawsuits Failure to generate enough revenue to cover costs	Ran out of cash and wasting the fund Losing credit against investors	Medium	High
<b>Technical</b>	Technical difficulties in developing and launching the product Failure of the product to meet user expectations Security breaches or cyberattacks	Decrease of profitability Client distrust Decreasing reputation	Medium	Medium
<b>Strategical</b>	Lack of flexibility and adaptability Unclear goals Inappropriate strategy Changes in user preferences or market conditions	Inability to be innovative Unrecognizing the ongoing market dynamics Inability to offer unique value for customers Decrease of profitability and customer relationship	Low	High
<b>Team</b>	Hiring unqualified members Disagreements or conflicts within the team High costs of key team members	Team conflicts Inability to reach milestones and goals	Low	Medium
<b>Unforeseen events</b>	Political changes Regulatory changes Negative publicity or social media backlash	Delays in achieving goals Delay in reaching milestone	Low	High

Figure 3.9: Risk Analysis

## 3.5 Development Life Cycle

### 3.5.1 Introduction

The part aims to realize the House of Quality (HOQ) to analyze the product that can generate soundtracks by analyzing videos using AI with its strengths and weaknesses in comparison to its existing competitors, ultimately aiming to identify areas for improvement and solidify its competitive edge in the market. The main customers of this product are music composers and content creators who want to enhance their video content with unique & royalty-free soundtracks while having the chance to control the generated soundtracks to achieve their desired results.

### 3.5.2 Customer Requirements

The customer requirements are represented as the “Voice of customers” and captured by different methods, such as personal interviews, surveys, focus groups, and other qualitative or product analysis techniques. It is essential to first identify them in order to provide a product that would satisfy their needs. To reach the final customer requirements that are used in the analysis, a sequence of steps must be followed. The starting point is the voice of the customer, taken as it is, and it is represented as a sentence in which the customer expresses his/her ideas about the product. We combined the answers and achieved complete sentences that define customer requirements.

- **Uniqueness & Copyright:**
  - "I want the generated music to be completely original and stand out from existing songs, but also be copyright-free or allow for easy clearance. This would be a game-changer for creators who need unique music without copyright headaches."
- **Simplicity & Good Looking:**
  - "I should be able to create high-quality, original music from a video with just a few clicks, all within a visually appealing and user-friendly environment."
- **Audio Length & Video Length:**
  - "The product should be flexible enough to handle videos and generate music of all lengths. Whether it's a short Instagram clip or a full-length documentary, I want the music to perfectly match the video's duration."
- **Achieving Desired Sound & Video Detection Capabilities:**

- "The aim should be the product to analyze the video content and automatically generate music that perfectly captures the mood, style, and genre.
- "Whether it's an energetic action scene or a peaceful nature documentary, the AI should be able to translate the visuals into a perfectly matching soundscape."

- **Database:**

- "While originality is important, having access to a curated database of high-quality sounds and samples would be a great addition. This way, I can further personalize the generated music by integrating these elements while still maintaining copyright-friendly options."

These combined statements showcase how different customer requirements can work together to create a more complete and user-friendly experience. In the end, the final customer requirements identified are the following ones. It can be noticed that they are grouped into macro-categories, based on their affinity:

- **USABILITY**

- User Experience (UX) Needs
  - \* Simplicity: The product should be easy to use for people with varying levels of technical expertise.
  - \* User Interface (UI): The interface should be visually appealing, well-organized, and responsive for a seamless user experience.
- Economic Needs
  - \* Price: The product's pricing model should be competitive and offer value for money to the target audience.

- **PERFORMANCE**

- Functional Needs
  - \* Uploaded Video Length: The product should support a range of video lengths that cater to user requirements.
  - \* File Format: The product should support commonly used video file & audio file formats for user convenience.
  - \* Generated Audio Length: The product should offer users control over the generated music track's length.
  - \* Video Detection Quality: The AI model should accurately analyze and detect elements within uploaded videos.

- \* Sound Generation Quality: The generated music should be high-quality, professional sounding, and free from technical glitches.
- Confidence Needs
  - \* Copyright: The product should ensure copyright protection of user-uploaded videos and avoid copyright infringement in generated music.
  - \* Data Privacy: The product should implement robust security measures to protect user data privacy for uploaded videos and generated content.
  - \* Uniqueness: The generated music should be unique and avoid sounding like derivative or repetitive compositions.
- Customization Needs
  - \* Modification Variety: The product should allow users to modify the generated music to some degree after creation. This involves adjusting tempo, mood, adding specific instruments, or fine-tuning certain aspects of the composition.
  - \* Genre & Instrument Variety: The product should offer users a wide range of music genres and instruments to choose from for music generation.

The priority of the customer requirements has been decided based on a survey in 30 surveyors participated. Two important measures are considered to understand the customers better. In the figure 3.10, the ratings each surveyor gave to each measure and their absolute and relative importance. The absolute importance is calculated by multiplying the number of votes for each rating and dividing it by the total number of responses and the Relative Importance is calculated by dividing each Absolute Importance by the total of absolute importance.

In the figure 3.11, shows us the absolute importance values for each customer requirements, and then the red dotted line represents the average while the green dotted line represents the 3rd quartile to let us understand which customer requirements are more crucial for the product. To summarize, the most important feature of our product would be to generate unique soundtracks while having a user-friendly interface. These results show that the product has to have a user-friendly interface with precise analysis & execution capabilities while giving our users to edit the generated songs with depth.

Customers Requirements	1	2	3	4	5	Absolute Importance	Relative Importance
Uniqueness	1	2	5	14	8	3.87	7.23%
Simplicity	0	3	6	15	6	3.80	7.10%
Design	0	0	11	8	11	4.00	7.48%
Economic	2	2	8	8	10	3.73	6.98%
Genre & Instrument Variety	3	3	7	12	5	3.43	6.42%
File Characteristics (video & audio length, format)	2	2	13	7	6	3.43	6.42%
Privacy and Security	3	8	9	7	3	2.97	5.55%
Modification Variety	1	3	8	12	6	3.63	6.79%
Short loading times	2	2	9	10	7	3.60	6.73%
Efficient Customer Service	9	9	10	1	1	2.20	4.11%
Built-in Plugins	2	7	9	7	5	3.20	5.98%
Internal AI assistance	0	3	9	8	10	3.83	7.17%
App Platform	0	6	3	7	14	3.97	7.41%
Precise Video Detection & Sound Generation	0	0	5	17	8	4.10	7.66%
Royalty Free Generations	0	3	10	9	8	3.73	6.98%

Figure 3.10: Customer Requirement Ratings and Importance Measures

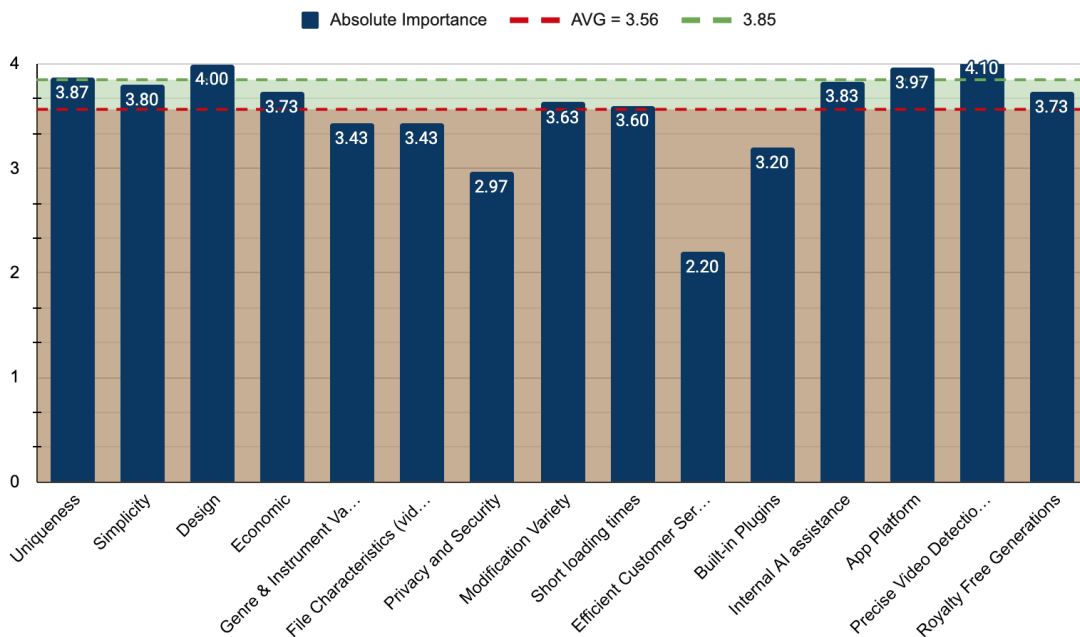


Figure 3.11: Customer Requirements vs. AVG of Absolute Importance

### 3.5.3 Competitive Benchmarking

The next step is competitive benchmarking. The goal is to identify the competitors, their strengths and weaknesses, and our products' position with respect to theirs.

The competitors are AIVA, Soundful, Studio by MatchTune and Dynascore.

First, a benchmarking analysis for each competitor can be seen from the figure 3.18. The same survey is used for the assessment and prioritization of CR to know which is the weight the competitors' customers give to the selected CR. Also in this case, the answers collected are 30.

- **AIVA:** For AIVA, in the figure 3.12, attributes like File Characteristics, Royalty Free Generations, and Short Loading Times emerged as top priorities for the customers. While they showed lower attention to the platform of the app, price and customer service.

Customers Requirements	1	2	3	4	5	Absolute Importance	Relative Importance
Uniqueness	0	2	5	15	8	3.97	8.76%
Simplicity	2	1	6	15	6	3.73	8.24%
Design	0	3	8	8	11	3.90	8.61%
Economic	4	7	8	8	3	2.97	6.55%
Genre & Instrument Variety	1	3	7	12	7	3.70	8.17%
File Characteristics (video & audio length, format)	0	0	0	4	26	4.87	10.74%
Privacy and Security	1	3	16	7	3	3.27	7.21%
Modification Variety	0	3	8	12	7	3.77	8.31%
Short loading times	0	0	2	10	18	4.53	10.01%
Efficient Customer Service	0	10	10	5	5	3.17	6.99%
Built-in Plugins	0	0	10	12	8	3.93	8.68%
Internal AI assistance	0	0	18	9	3	3.50	7.73%
App Platform	10	5	5	5	5	2.67	5.89%
Precise Video Detection & Sound Generation	1	1	0	8	20	4.50	9.93%
Royalty Free Generations	0	0	0	11	19	4.63	10.23%

Figure 3.12: AIVA CR

- **Studio by MatchTune:** For MatchTune, in the figure 3.13, attributes like Precise Video Detection & Sound Generation, Royalty Free Generations and File Characteristics are top priorities. While having built-in plugins, efficient customer service and modification variety have lower priority.
- **Dynascore:** For Dynascore, in the figure 3.14, high scores in design, simplicity, and internal AI assistance indicate these are critical areas where customers expect significant functionality and value. Low scores for file characteristics and royalty-free generations suggest these areas are less critical to the current customer base for the purpose of application.



Customers Requirements	1	2	3	4	5	Absolute Importance	Relative Importance
Uniqueness	4	5	5	6	8	3.10	7.98%
Simplicity	2	1	6	15	6	3.73	9.61%
Design	0	3	8	8	11	3.90	10.04%
Economic	4	7	8	8	3	2.97	7.64%
Genre & Instrument Variety	3	3	7	10	7	3.50	9.01%
File Characteristics (video & audio length, format)	1	0	5	12	12	4.13	10.64%
Privacy and Security	1	13	6	7	3	2.93	7.55%
Modification Variety	4	12	8	3	3	2.63	6.78%
Short loading times	7	3	5	10	5	3.10	7.98%
Efficient Customer Service	5	10	9	2	4	2.67	6.87%
Built-in Plugins	6	6	15	2	1	2.53	6.52%
Internal AI assistance	0	2	13	9	6	3.63	9.36%
App Platform	4	11	5	5	5	2.87	7.38%
Precise Video Detection & Sound Generation	1	1	0	6	22	4.57	11.76%
Royalty Free Generations	0	0	0	15	15	4.50	11.59%

Figure 3.13: Studio by MatchTune CR

Customers Requirements	1	2	3	4	5	Absolute Importance	Relative Importance
Uniqueness	1	7	10	10	2	3.17	8.70%
Simplicity	0	3	2	15	10	4.07	11.17%
Design	0	0	3	12	15	4.40	12.09%
Economic	1	11	8	10	1	3.07	8.42%
Genre & Instrument Variety	3	10	9	8	0	2.73	7.51%
File Characteristics (video & audio length, format)	15	14	0	0	1	1.60	4.40%
Privacy and Security	10	10	5	5	0	2.17	5.95%
Modification Variety	10	18	2	0	0	1.73	4.76%
Short loading times	3	2	10	10	5	3.40	9.34%
Efficient Customer Service	3	2	10	10	5	3.40	9.34%
Built-in Plugins	6	6	6	6	6	3.00	8.24%
Internal AI assistance	2	2	10	6	10	3.67	10.07%
App Platform	6	5	5	9	5	3.07	8.42%
Precise Video Detection & Sound Generation	5	7	5	8	5	3.03	8.33%
Royalty Free Generations	15	11	4	0	0	1.63	4.49%

Figure 3.14: Dynascore CR

- **Soundful:** For Soundful, in the figure 3.15, Soundful places significant emphasis on Genre & Instrument Variety (4.07), Efficient Customer Service (3.77), and Internal AI Assistance (3.77), showcasing a strong commitment to diversity in music options and customer support. With Royalty Free Generations (4.03) also scoring highly, Soundful aims to cater to users seeking hassle-free music creation without legal concerns. The application balances technical attributes with user-friendly features, as seen in its high scores for Simplicity (4.13) and Uniqueness (3.77).

According to the characteristics of the competitors, the positioning of Soundtrackr can be seen from the figures 3.16 & 3.17.

Customer Requirements	1	2	3	4	5	Absolute Importance	Relative Importance
Uniqueness	2	3	5	10	10	3.77	8.50%
Simplicity	1	2	4	8	15	4.13	9.33%
Design	2	4	6	10	8	3.60	8.13%
Economic	3	5	8	6	8	3.37	7.60%
Genre & Instrument Variety	1	2	4	10	13	4.07	9.18%
File Characteristics (video & audio length, format)	3	3	5	8	11	3.70	8.35%
Privacy and Security	2	3	7	9	9	3.67	8.28%
Modification Variety	1	3	6	10	10	3.83	8.65%
Short loading times	4	5	7	8	6	3.23	7.30%
Efficient Customer Service	2	3	6	8	11	3.77	8.50%
Built-in Plugins	3	4	7	10	6	3.40	7.67%
Internal AI assistance	2	3	5	10	10	3.77	8.50%
App Platform	1	4	6	10	9	3.73	8.43%
Precise Video Detection & Sound Generation	3	5	6	7	9	3.47	7.83%
Royalty Free Generations	1	2	5	9	13	4.03	9.10%

Figure 3.15: Soundful CR

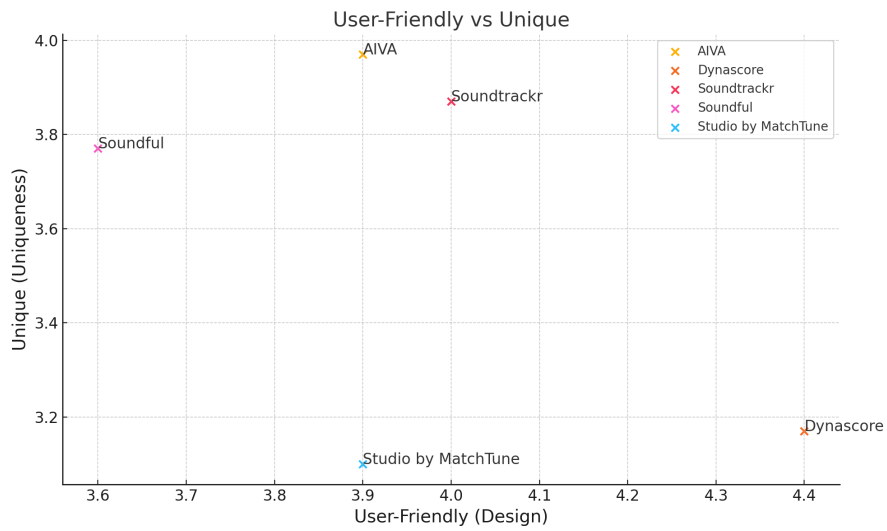
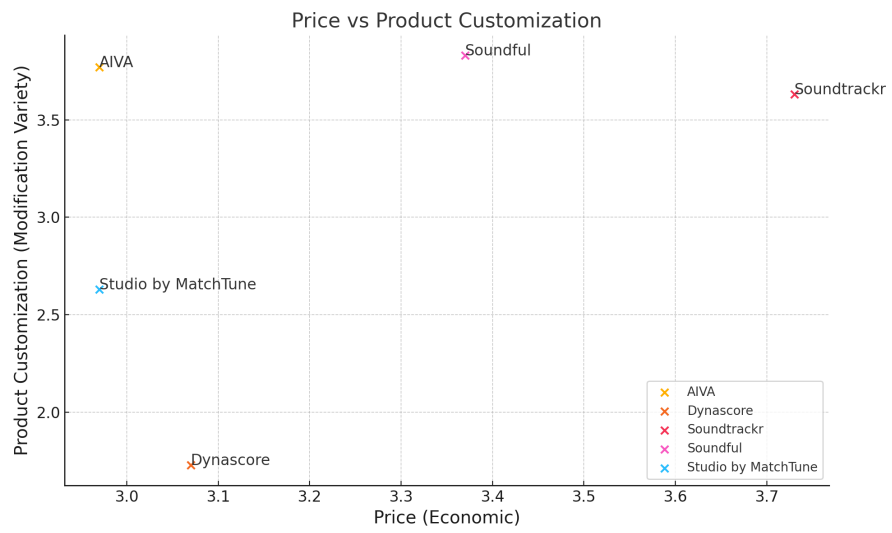


Figure 3.16: User - Friendly vs Unique



**Figure 3.17:** Price vs Product Customization

Competitor	Uniqueness	Simplicity	Design	Economic	Genre & Instrument Variety	File Characteristics (video & audio length, format)	Privacy and Security	Modification Variety	Short loading times	Efficient Customer Service	Built-in Plugins	Internal AI assistance	App Platform	Precise Video Detection & Sound Generation	Royalty Free Generations
AVA	Full control via software with editing capabilities and wide variety of options for generation.	Easy to use with drag-and-drop interface.	Modern and clean design.	Subscription-based pricing.	Limited to classical and cinematic genres & instruments.	Audio: MP3, MIDI, and high-quality WAV files. Length: Offers some control over song length.	Sufficient	Highly customizable	Sufficient	Sufficient	Accurate sound generation according to the options selected by user.	Needed	Software and webpage is enough.	Decent	Needed
Soundful	Curated music library, not strictly AI-generated. Limited originality compared to AI composition.	Simple interface with search functionality.	Basic and functional design.	Subscription-based pricing.	Wide variety of genres & instruments in their curated library.	Audio: MP3, WAV, and STEM file formats. Length: Offers some control over song length.	Sufficient	Limited customization options	Sufficient	Sufficient	Limited sound generation to directly match video content with music selection.	Not needed	Mobile app would be nice.	Ok	Needed
Studio by Matchtune	Analyses video and presents with several matching tracks that have already been edited to match video length and sync with any cuts.	Straightforward interface for genre selection	Basic and utilitarian design	Freemium with paid plans for advanced features	Limited to offered genres & instruments	For video: MP4, MOV. For audio: MP3, WAV. Length: Supports most video lengths.	Sufficient	Limited customization options	Sufficient	Sufficient	Accurate sound generation but in the end depends on the user. Video detection quality is high.	Not needed	Included in all platforms needed	Decent	Needed
Dynascore	Video-synchro nized music editing software integration	Interface geared towards composers, steeper learning curve	Functional design	Subscription-based pricing	Variety of genres & instruments	All. Length: Depends on the video uploaded.	Sufficient	Limited customization options	Sufficient	Sufficient	Limited sound generation to directly match video content with music selection. Video analysis is automatic and editable.	Not needed	Included in all platforms needed	Ok	Needed

Figure 3.18: Competitor Benchmarking

### 3.5.4 Technical Characteristics

These technical specifications define the operational parameters and performance metrics that govern the functionality and user experience of our product. Each characteristic, ranging from input/output formats and processing capabilities to user customization options and system dimensions—is crucial for tailoring the software to meet diverse user needs and industry standards.

- **Video Input Format:** The variety of video file formats that the product can accept for analysis.
  - Unit of measurement: # formats
  - Direction of Improvement: Increase the number of supported formats to cater to a wider user base.
- **Video Length:** The length of video file that the product can accept for analysis.
  - Unit of measurement: minutes [min]
  - Direction of Improvement: Increase the maximum video length to handle longer user content.
- **Audio Output Formats:** The variety of audio file formats that the product generates.
  - Unit of measurement: # formats
  - Direction of Improvement: Offer a wider variety of output formats for user preference and flexibility.
- **Music Length:** The ability for users to specify the desired length of the generated music track.
  - Unit of measurement: minute [min]
  - Direction of Improvement: Allow for more granular control over music track length for precise user needs.
- **Processing Time:** The time it takes to analyze a video and generate music.
  - Unit of measurement: seconds [sec]
  - Direction of Improvement: Reduce processing time to improve user experience and efficiency.

- **DB Capacity:** The capacity of DB.
  - Unit of measurement: # data
  - Direction of Improvement: Increasing the capacity to have better results for video analysis and sound generation.
- **Dimension of Software:** The dimension of software.
  - Unit of measurement: megabytes [mb]
  - Direction of Improvement: Decreasing the dimension will attract new customers.
- **Music Genre Options:** The range of musical genres users can choose from for music generation.
  - Unit of measurement: # genre
  - Direction of Improvement: Increase the genres.
- **Customization Options:** The degree to which users can modify the generated music after creation.
  - Unit of measurement: # options
  - Direction of Improvement: Increase the customization to enhance creativity even more.
- **Software Theme:** The visual appearance, organization, and ease of use of the product's interface.
  - Unit of measurement: type
  - Direction of Improvement: Increase in order to enhance simplicity and appearance.
- **Communication System Efficiency:** The efficiency between users and customer service.
  - Unit of measurement: Integer [int]
  - Direction of Improvement: Increasing the efficiency to help new users.
- **Feedback Collection:** The variety of operating systems
  - Unit of measurement: # feedback
  - Direction of Improvement: Increasing the feedback collection will increase the precision of video analysis and sound generation.

- **Licensing & Privacy Policy:** The measures in place to protect user data privacy and uploaded videos.
  - Unit of measurement: Integer [int]
  - Direction of Improvement: Increase the efficiency for more trustworthy software.
- **ML Models:** The specific functionalities offered for video analysis (e.g., scene detection, object recognition, facial recognition) and sound generation.
  - Unit of measurement: models
  - Direction of Improvement: Increasing the models will increase the depth of the software and the uniqueness of sound generation.
- **AI Libraries:** AI libraries for increased functionalities
  - Unit of measurement: libraries
  - Direction of Improvement: Increasing the libraries will increase the depth of the software and the uniqueness of sound generation.

### 3.5.5 Relationship Matrix

The relationship matrix is defined by the cross-functional team to identify which are the relationships between CR and TC. It identifies the main correlations among them, defining the core of the House of Quality. In the analyzed case, three degrees of correlations have been identified:

- “●” represents a strong correlation. It is substituted with a value of “9” points to compute the absolute importance and weight of the TCs.
- “○” represents a medium correlation. It is substituted with a value of “3” points to compute the absolute importance and weight of the TCs.
- “S” represents a weak correlation. It is substituted with a value of “1” points to compute the absolute importance and weight of the TCs.

The final matrix obtained in the analyzed case is displayed in the following figure 3.19. In particular, the direction of improvement, the unit of measure, and the absolute weight have been inserted in the same table.

Customer Requirements / Technical Characteristics		Video Input Format	Video Length	Audio Output Formats	Music Length	Processing Time	DB Capacity	Dimension of Software	Music Genre Options	Customization Options	Software Theme	Communication System Efficiency	Feedback Collection	Licensing & Privacy Policy	ML Models	AI Libraries
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Uniqueness				o		•		•				S		•	•
2	Simplicity					o			o	o	o	S			o	S
3	Design							S			•					
4	Economic		•		•	•	•	o	o	•	S				o	o
5	Genre & Instrument Variety				•	•	•	S	•	•					o	o
6	File Characteristics (video & audio length, format)	•	•	•	•	•	o	S							S	S
7	Privacy and Security													•		
8	Modification Variety					•	•	S	o	•					o	o
9	Short loading times		•		•	•	•	o							•	•
10	Efficient Customer Service											•				
11	Built-in Plugins					•	o	•		o					S	o
12	Internal AI assistance						•		o	o			•		•	•
13	App Platform					•	•				o					
14	Precise Video Detection & Sound Generation		o			•	•	o	o	o			o		•	•
15	Royalty Free Generations					o	•		S	S			o	•	S	S
<i>Direction of improvement</i>		↑	↑	↑	↑	↓	↑	↓	↑	↑	↑	↑	↑	↑	↑	↑
<i>Unit of measurement</i>		# formats	min	format	min	sec	# data	mb	# genre	# option	type	Int	# feedback	Int	models	libraries
<i>Absolute weight</i>		0.50	1.55	0.50	1.73	3.00	4.31	1.34	2.00	3.11	0.62	0.31	0.95	0.34	3.00	3.04
<i>Relative weight</i>		1.9%	5.9%	1.9%	6.6%	11.4%	16.4%	5.1%	7.6%	11.8%	2.3%	1.2%	3.6%	1.3%	11.4%	11.6%

Figure 3.19: Relationship Matrix

### 3.5.6 House of Quality

Grouping all the previous sections in a single graph we will obtain the House of Quality (HoQ) (figure 3.20), displayed below. Here, it is possible to notice the Customer Requirements with their degree of importance on the left-hand side. Then, in the right-hand part of the graph, the Benchmarking of the CRs is found, with the relative improvement that the team has concerning the existing models. In the central part is it possible to find the Relationship Matrix, with the Technical Characteristics in the upper part and the relative weight in the bottom. In the upper part of the house, the Relationships among TCs are displayed in the so-called “roof”. In the end, in the bottom part, it is possible to find the Benchmarking of the TCs, in order to show the improvement that the current product is bringing with respect to the existing ones.



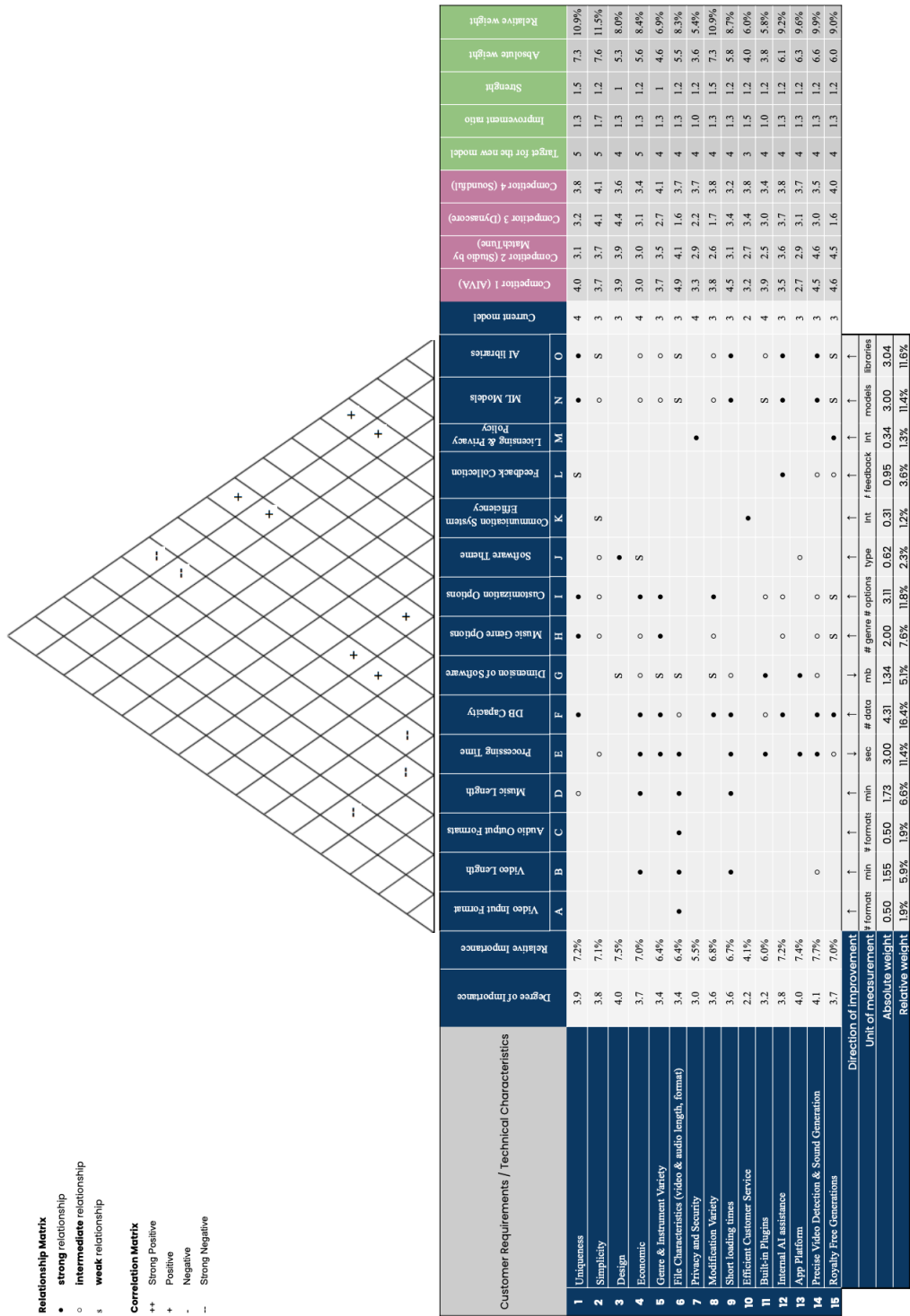


Figure 3.20: House of Quality

### 3.5.7 Lyman’s Normalization

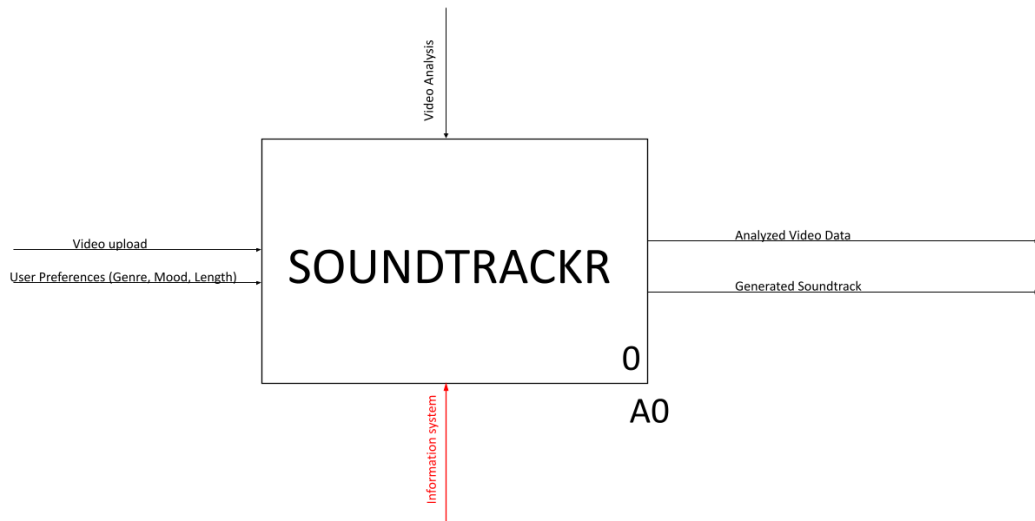
The Lyman’s Normalization 3.21 is a tool used for correcting the ISM’s propensity to be dependent on the number of sub-characteristics that refer to technical ones. By normalizing the coefficients in the relationship matrix, the TCs related to the low weight CRs will have lower importance. This table shows that the ranking position has changed mainly for the characteristics that have direct impact on users’ software interaction. While DB capacity and processing time are the most important characteristics with or without the normalization.

Technical Characteristics		Video Input Format		Video Length		Audio Output Formats		Music Length		Processing Time		DB Capacity		Dimension of Software		Music Genre Options		Customization Options		Software Theme		Communication System Efficiency		Feedback Collection		Licensing & Privacy Policy		ML Models		AI Libraries	
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB		
Technical relative importance	Normal Ranking	13	8	13	9	2	1	7	6	5	10	15	11	12	3	4															
	no Lyman	58%	204%	58%	203%	471%	541%	212%	237%	337%	118%	44%	116%	113%	360%	358%															
	Lyman	1%	4%	1%	4%	13%	13%	8%	6%	9%	9%	5%	3%	8%	9%	8%															
	Lyman's Ranking	14	11	14	12	2	1	8	9	4	3	10	13	7	5	6															
Relative weight	Normal Ranking	13	8	13	9	2	1	7	6	5	12	15	10	11	4	3															
	no Lyman	50%	188%	50%	173%	474%	619%	213%	243%	354%	77%	31%	157%	124%	408%	412%															
	Lyman	1%	3%	1%	3%	13%	16%	7%	6%	10%	6%	3%	4%	7%	10%	10%															
	Lyman's Ranking	14	11	14	12	2	1	6	8	3	9	13	10	7	4	5															

Figure 3.21: Lyman’s Normalization

### 3.5.8 IDEF0

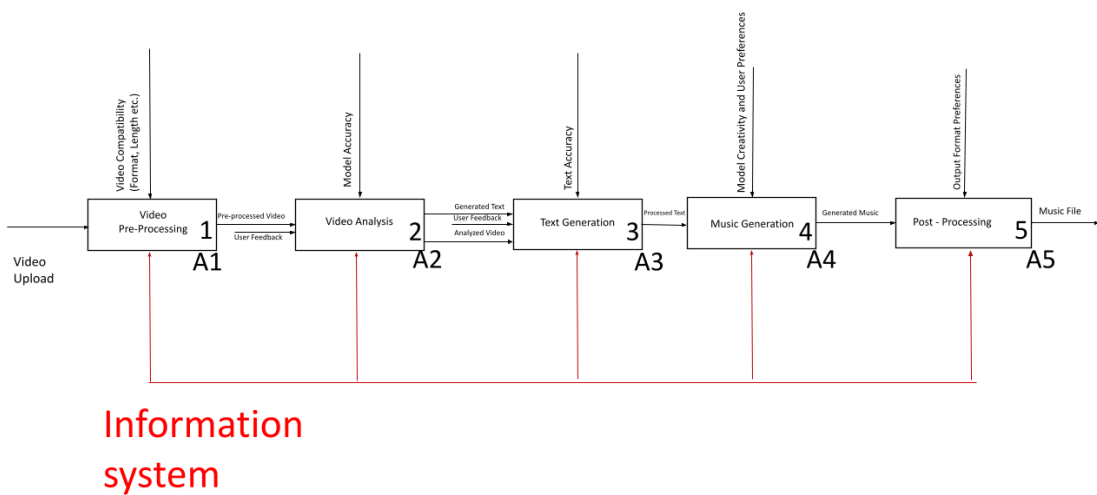
In the IDEF0 are represented the main processes of the web-app. Thanks to this tool it is possible to clarify everything connected to these processes such as inputs, outputs, and resources. The IDEF0 analysis is divided into different levels. In this analysis, it is started from a general view and then moved the focus to more specific processes. In the figure 3.22, it can be seen the general view of the product. The inputs are the video uploaded and the user preferences for the soundtrack generation while the outputs are analyzed video data and generated soundtrack. The DB and ML models are inside the information systems and the main functionality of the product starts with the video analysis.



A0	Soundtrackr
----	-------------

Figure 3.22: IDEF0: General View

In the figure 3.23, a more detailed view can be seen and the the product is separated to 5 processes.

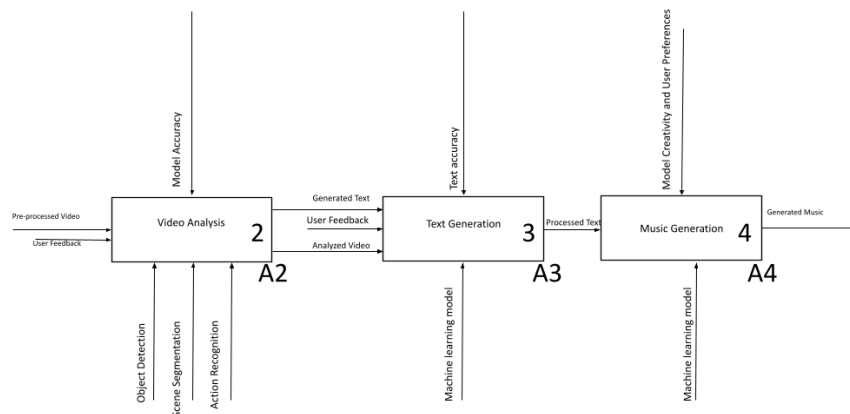


A0	Soundtrackr	Number 1
----	-------------	----------

Figure 3.23: IDEF0: Soundtrackr Overview

- **A1, Video Pre-processing:** The system checks the format, length, and other metrics to start the processing.
- **A2, Video Analysis:** It takes the pre-processed video and starts the analysis of it by the ML models. The analysis will be conducted combining user feedback and the ML models' work. In the end, it will generate a text-based description of the video.
- **A3, Text Generation:** In this part, the accuracy of the text will be analyzed and the output is the processed text.
- **A4, Music Generation:** The system takes the processed text and using ML models and user preferences, it will generate soundtracks.
- **A5, Post - Processing:** The system will ask the user what formatting they want and then provide it.

In the Figure 3.24 shows a more detailed version of the main processes of the product that is crucial for the software. For video analysis, different types of ML models will be used to analyze videos and retrieve information about them.

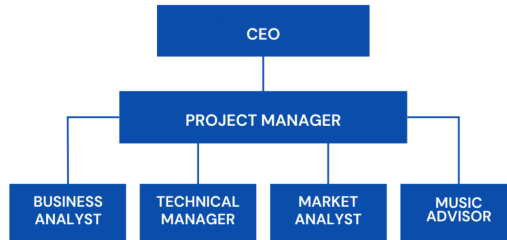


A0 - 1	Foundation	Number 2
--------	------------	----------

Figure 3.24: IDEF0: Detailed View

## 3.6 Organizational Plan

Soundtrackr's organizational structure is designed to streamline operations and promote effective management across different areas of our business. It will be organized as follows 3.25:



**Figure 3.25:** Organizational Chart

- **CEO** - Responsible for managing a company's overall operations. This may include delegating and directing agendas, driving profitability, managing company organizational structure, strategy, and communicating with the board
- **Project Manager** - Responsible for workflow management and organization. Deadlines and deliverables check.
- **Technical Manager** - Responsible for the development, implementation, and maintenance of the main technologies.
- **Team Manager** - Set targets, implement guidelines, and assist with any issues the team may have. Ensure the proper conduct of meetings and harmony in the team.
- **Market Analyst** - Provides actionable overall market and customer insights to address key strategic decisions. Collect and analyze qualitative data, trends, and information on competitors.
- **Business Analyst** - Responsible for understanding and assessing the changing needs of the project. Coordinate with the team to optimize results.
- **Software Engineer, Full Stack Developer, R&D Engineers** - These technical roles are crucial for developing and innovating our product offerings.

In the future, with the growth of the product and its service, it will be necessary to reconfigure the organizational structure following the recruitment of key figures for the improvement of the platform and strategy.

## **3.7 Project Plan**

For the completion of the start-up phase, Soundtrackr intends to complete the activities that are reported in the project plan shown in 3.27. These activities start with the presentation of the business plan, which sets milestones regarding the most important goals to be achieved. The Business Plan Preparation phase spans almost a full year. This stage involves the comprehensive planning and structuring of the project, covering all aspects from market analysis to operational and financial planning. This foundational step is crucial for setting clear objectives and strategies for the entire project lifecycle. The App Development phase follows the initial MVP development and lasts approximately three and a half months. This phase likely focuses on building a more robust and fully-featured version of the app, incorporating feedback from the MVP stage and preparing the app for a broader market release. The MVP development is designed to quickly get a workable product to market to test assumptions and gather user feedback. This phase includes the development and testing of both the app and its accompanying website, concluding with the launch of the MVP to gauge initial user engagement and feedback. Activities during the first year post-launch are critical for the project's early success and include ongoing development, marketing efforts, user support, and iterative improvements based on user feedback. This period is essential for establishing a strong market presence and beginning the scaling process. These phases collectively represent a well-structured approach to developing and launching a new software product, with a clear emphasis on planning, testing, refining ofining the product based on market feedback. This strategic planning ensures that the project remains focused and adaptable to changes in market conditions or user preferences.

## Roadmap

The roadmap of Soundtrackr outlines the key phases that the company intends to overcome.

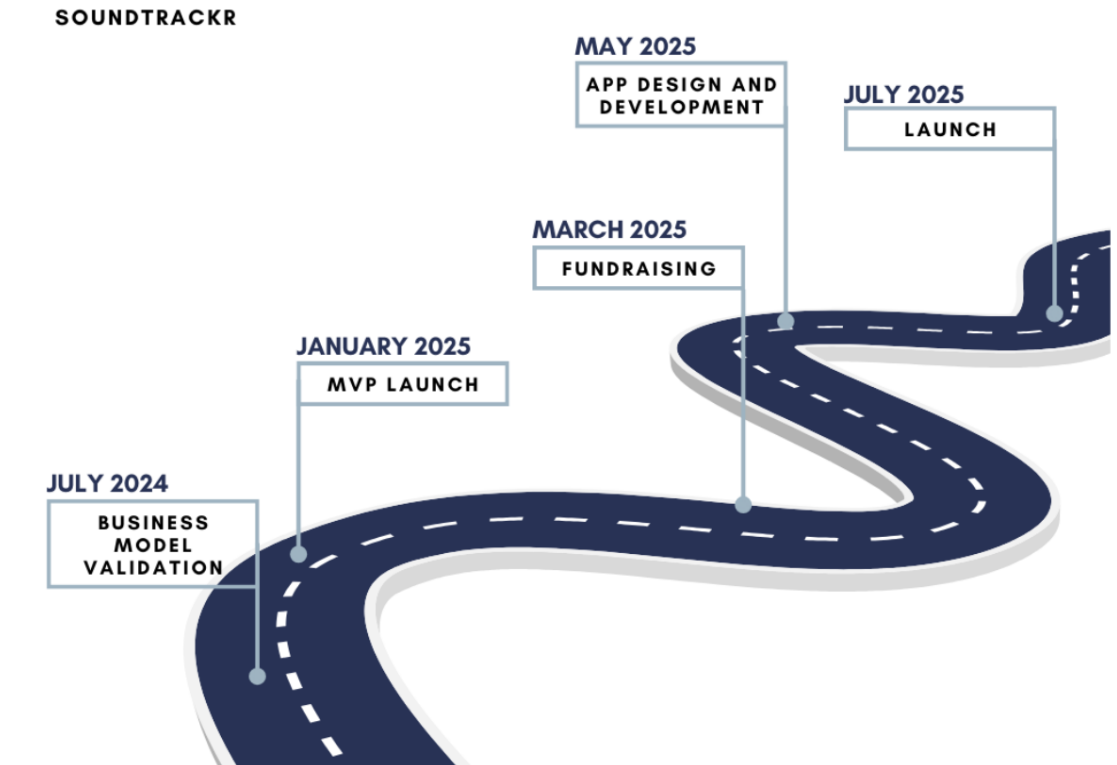


Figure 3.26: Roadmap of Soundtrackr

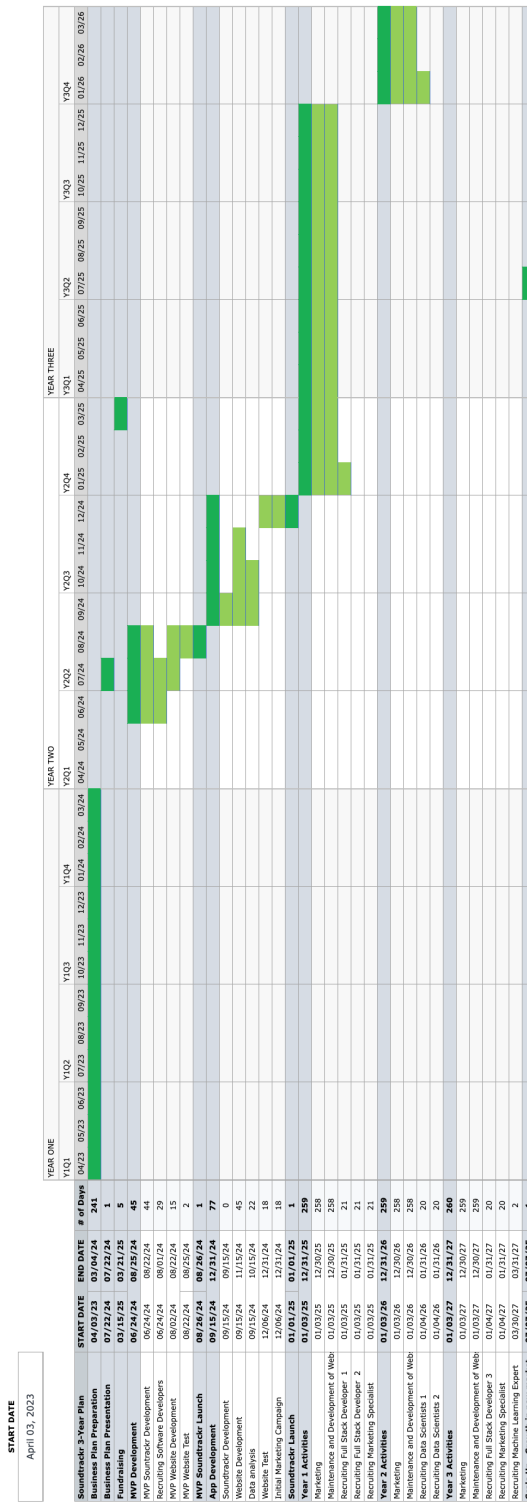


Figure 3.27: Gantt Chart of Soundtrackr



# Chapter 4

## Conclusion

### 4.1 AI Overview

The use of artificial intelligence techniques for the creation of soundtracks from videos holds significant potential in the media industry. The integration of AI-powered soundtrack creation can lead to a more seamless and emotionally engaging storytelling experience by dynamically generating music that reflects the evolving emotions and themes within the visual narrative. However, it is crucial for businesses in this field to consider changing consumer preferences, ethical and cultural sensitivities, and the evolving technological landscape. The demand for AI in the music industry is on the rise, and the rapid growth of the video content market, particularly in social media, presents opportunities for innovative AI-driven solutions. The successful development and implementation of AI tools for music generation and video analysis can address the needs of customers for quick, easy, and affordable soundtrack creation, while also providing them with greater control and inspiration. Overall, the careful consideration of technical, ethical, and artistic implications is essential in harnessing the full potential of AI for soundtrack generation from videos.

## **4.2 Business Model Conclusion**

With the business model, it is concluded that this type of a startup will be profitable in the long term due to the freshness of artificial intelligence usage in the music generation industry since it requires complex algorithms and huge databases to create soundtracks but also to guide users during their production phases. Also, we need to consider the prejudices that AI generated music may have. These prejudices will fade with the new developments of AI. AI will undoubtedly improve productions for a variety of customer segments, but in the short term, it would be more advantageous to concentrate on content creators because of their rapid video consumption and lower need for sophisticated musical compositions. This is why, when thinking of a project like this, it is important to follow the trends, upgrade the product when it is possible, focus heavily on marketing when the product is ready, and reduce unnecessary costs to reduce risks.

In conclusion, this project requires patience and commitment to stay relevant and have profitability.

# Appendix A

## Survey Questions

### Survey Questions

No.	Question
1	<p data-bbox="762 925 1102 958">What is your age group?</p> <ul data-bbox="411 992 632 1406" style="list-style-type: none"><li data-bbox="411 992 571 1025">• Under 18</li><li data-bbox="411 1055 520 1088">• 18-24</li><li data-bbox="411 1117 520 1151">• 25-34</li><li data-bbox="411 1180 520 1214">• 35-44</li><li data-bbox="411 1243 520 1276">• 45-54</li><li data-bbox="411 1305 520 1339">• 55-64</li><li data-bbox="411 1368 632 1402">• 65 and above</li></ul>
2	<p data-bbox="783 1451 1082 1485">What is your gender?</p> <ul data-bbox="411 1518 679 1738" style="list-style-type: none"><li data-bbox="411 1518 512 1552">• Male</li><li data-bbox="411 1581 544 1615">• Female</li><li data-bbox="411 1644 603 1677">• Non-binary</li><li data-bbox="411 1706 679 1740">• Prefer not to say</li></ul>

Continued on next page

**Table A.1 – continued from previous page**

No.	Question
3	<p style="text-align: center;">What is your primary occupation?</p> <ul style="list-style-type: none"> <li>• Content creator</li> <li>• Video editor</li> <li>• Musician</li> <li>• Other (please specify)</li> </ul>
4	<p style="text-align: center;">How important is it for you that the application is easy to use, regardless of your technical expertise?</p> <ul style="list-style-type: none"> <li>• 5 - Very important</li> <li>• 4 - Important</li> <li>• 3 - Moderately important</li> <li>• 2 - Slightly important</li> <li>• 1 - Not important at all</li> </ul>
5	<p style="text-align: center;">How would you rate the visual appeal of an application's user interface on a scale of 1 to 5?</p> <ul style="list-style-type: none"> <li>• 5 - Very attractive</li> <li>• 4 - Attractive</li> <li>• 3 - Neutral</li> <li>• 2 - Unattractive</li> <li>• 1 - Very unattractive</li> </ul>
Continued on next page	

**Table A.1 – continued from previous page**

No.	Question
6	<p style="text-align: center;">How important is responsiveness and quick performance for you in a video-to-music generation application?</p> <ul style="list-style-type: none"> <li>• 5 - Very important</li> <li>• 4 - Important</li> <li>• 3 - Moderately important</li> <li>• 2 - Slightly important</li> <li>• 1 - Not important at all</li> </ul>
7	<p style="text-align: center;">How intuitive do you find the user interface of video and music editing software you currently use?</p> <ul style="list-style-type: none"> <li>• 5 - Very intuitive</li> <li>• 4 - Intuitive</li> <li>• 3 - Moderately intuitive</li> <li>• 2 - Slightly intuitive</li> <li>• 1 - Not intuitive at all</li> </ul>
8	<p style="text-align: center;">How would you evaluate the importance of the application's pricing in your decision to use it?</p> <ul style="list-style-type: none"> <li>• 5 - Extremely important</li> <li>• 4 - Very important</li> <li>• 3 - Moderately important</li> <li>• 2 - Slightly important</li> <li>• 1 - Not important at all</li> </ul>
Continued on next page	

**Table A.1 – continued from previous page**

No.	Question
9	<p style="text-align: center;">What pricing model do you prefer for such applications?</p> <ul style="list-style-type: none"> <li>• One-time purchase</li> <li>• Subscription-based</li> <li>• Freemium</li> <li>• Pay-per-use</li> <li>• Other (please specify)</li> </ul>
10	<p style="text-align: center;">How critical is it that the application supports a wide range of video lengths (e.g., short clips to full-length documentaries)?</p> <ul style="list-style-type: none"> <li>• 5 - Very critical</li> <li>• 4 - Critical</li> <li>• 3 - Moderately critical</li> <li>• 2 - Slightly critical</li> <li>• 1 - Not critical at all</li> </ul>
11	<p style="text-align: center;">Which video file formats do you most commonly use? (Select all that apply)</p> <ul style="list-style-type: none"> <li>• MP4</li> <li>• AVI</li> <li>• MOV</li> <li>• WMV</li> <li>• Other (please specify)</li> </ul>
Continued on next page	

**Table A.1 – continued from previous page**

No.	Question
12	<p>How important is it that the generated music length matches the video length?</p> <ul style="list-style-type: none"> <li>• 5 - Extremely important</li> <li>• 4 - Very important</li> <li>• 3 - Moderately important</li> <li>• 2 - Slightly important</li> <li>• 1 - Not important at all</li> </ul>
13	<p>How would you rate the importance of accurate video content analysis for generating matching music?</p> <ul style="list-style-type: none"> <li>• 5 - Extremely important</li> <li>• 4 - Very important</li> <li>• 3 - Moderately important</li> <li>• 2 - Slightly important</li> <li>• 1 - Not important at all</li> </ul>
14	<p>How important is it that the generated music is high-quality and professional sounding?</p> <ul style="list-style-type: none"> <li>• 5 - Extremely important</li> <li>• 4 - Very important</li> <li>• 3 - Moderately important</li> <li>• 2 - Slightly important</li> <li>• 1 - Not important at all</li> </ul>
Continued on next page	

**Table A.1 – continued from previous page**

No.	Question
15	<p>How important is it for the application to support common audio file formats?</p> <ul style="list-style-type: none"> <li>• 5 - Extremely important</li> <li>• 4 - Very important</li> <li>• 3 - Moderately important</li> <li>• 2 - Slightly important</li> <li>• 1 - Not important at all</li> </ul>
16	<p>How crucial is it for you that the generated music is copyright-free or offers easy clearance?</p> <ul style="list-style-type: none"> <li>• 5 - Very crucial</li> <li>• 4 - Crucial</li> <li>• 3 - Moderately crucial</li> <li>• 2 - Slightly crucial</li> <li>• 1 - Not crucial at all</li> </ul>
17	<p>How important is data privacy and security for your uploaded videos and generated music?</p> <ul style="list-style-type: none"> <li>• 5 - Extremely important</li> <li>• 4 - Very important</li> <li>• 3 - Moderately important</li> <li>• 2 - Slightly important</li> <li>• 1 - Not important at all</li> </ul>
Continued on next page	



**Table A.1 – continued from previous page**

No.	Question
18	<p>How important is it for you that the generated music is unique and not repetitive?</p> <ul style="list-style-type: none"> <li>• 5 - Extremely important</li> <li>• 4 - Very important</li> <li>• 3 - Moderately important</li> <li>• 2 - Slightly important</li> <li>• 1 - Not important at all</li> </ul>
19	<p>How important is it for you to have the ability to modify the generated music (e.g., adjusting tempo, mood, adding instruments)?</p> <ul style="list-style-type: none"> <li>• 5 - Extremely important</li> <li>• 4 - Very important</li> <li>• 3 - Moderately important</li> <li>• 2 - Slightly important</li> <li>• 1 - Not important at all</li> </ul>
20	<p>How valuable would you find access to a curated database of high-quality sounds and samples for personalizing generated music?</p> <ul style="list-style-type: none"> <li>• 5- Very valuable</li> <li>• 4 - Valuable</li> <li>• 3 - Moderately valuable</li> <li>• 2 - Slightly valuable</li> <li>• 1 - Not valuable at all</li> </ul>
Continued on next page	

**Table A.1 – continued from previous page**

No.	Question
21	<p style="text-align: center;">How important is it for the application to offer a wide range of music genres and instruments for music generation?</p> <ul style="list-style-type: none"> <li>• 5 - Extremely important</li> <li>• 4 - Very important</li> <li>• 3 - Moderately important</li> <li>• 2 - Slightly important</li> <li>• 1 - Not important at all</li> </ul>
22	<p style="text-align: center;">How frequently do you modify the music or soundtracks in your video projects?</p> <ul style="list-style-type: none"> <li>• Always</li> <li>• Often</li> <li>• Sometimes</li> <li>• Rarely</li> <li>• Never</li> </ul>
23	<p>What additional features would you find most valuable in a video-to-music generation application? (Open-ended)</p>
24	<p>Examine the questions above and pinpoint the main requirements focus for the application with one word and a short phrase. (Open-ended)</p>

# Bibliography

- [1] Goldmedia. «AI and Music: Market Development of AI in the Music Sector and Impact on Music Authors and Creators». In: (2023). URL: <https://www.goldmedia.com/aktuelles/info/article/ai-and-music-market-development-of-ai-in-the-music-sector-and-impact-on-music-authors-and-creators/> (cit. on p. 1).
- [2] market.us. «Global Generative AI in Music Market By Component (Software and Services), By Type (GANs, AR-CNNs, Transformer-based Models), By Application, By Region and Companies - Industry Segment Outlook, Market Assessment, Competition Scenario, Trends, and Forecast 2023-2032». In: (Nov. 2023). URL: <https://market.us/report/generative-ai-in-music-market/> (cit. on pp. 2, 24).
- [3] Sturm, B.L.T.; Iglesias, M.; Ben-Tal, O.; Miron, M.; Gómez, E. «Artificial Intelligence and Music: Open Questions of Copyright Law and Engineering Praxis.» In: 8 (Sept. 2019), p. 115 (cit. on p. 2).
- [4] Bedroom Producers Blog. *A.I. Music Survey*. May 2023. URL: <https://bedroomproducersblog.com/2023/05/30/ai-music-survey/> (cit. on pp. 2, 25).
- [5] IBM. *Machine Learning*. URL: <https://www.ibm.com/topics/machine-learning> (cit. on p. 5).
- [6] LeCun, Y., Bengio, Y., & Hinton, G. E. «Deep Learning. Nature». In: 521 (2015), pp. 436–444 (cit. on p. 5).
- [7] Cevallos, A., Latorre, L., Alicandro, G., Wanner, Z., Cerrato, I., Zarate, J. D., Alvarez, J., Villacreses, K., Pfeifer, M., Gutierrez, M., Villanueva, V., Rivera-Fournier, A., Riobó, A., Pombo, C., Puerto, F., & Rodriguez Breuning, J. «Tech Report Generative AI». In: (May 2023). URL: <https://publications.iadb.org/en/tech-report-generative-ai> (cit. on p. 6).
- [8] Turing. *Deep Learning vs Machine Learning: The Ultimate Battle*. URL: <https://www.turing.com/kb/ultimate-battle-between-deep-learning-and-machine-learning> (cit. on p. 6).

- [9] International Journal of Robotics and Control Systems. «Understanding of Convolutional Neural Network (CNN): A Review». In: 2(4) (2023), pp. 739–748 (cit. on p. 6).
- [10] J. Wang and X. Hu. «Convolutional Neural Networks With Gated Recurrent Connections». In: 44 (2022), pp. 3421–3435 (cit. on p. 6).
- [11] Fischer, T., and Krauss, C. «Deep learning with long short-term memory networks for financial market predictions». In: 270 (2017), pp. 654–669 (cit. on p. 7).
- [12] Google. *Overview of GAN Structure*. URL: [https://developers.google.com/machine-learning/gan/gan\\_structure](https://developers.google.com/machine-learning/gan/gan_structure) (cit. on p. 7).
- [13] McKinsey & Company. *The economic potential of generative AI: The next productivity frontier*. URL: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#industry-impacts> (cit. on pp. 8, 24).
- [14] A. Magar, A. Acharya, S. Bothe, H. Bihani, and T. Desai. «Automatic Music Generation». In: (2023) (cit. on p. 8).
- [15] K. Choi, G. Fazekas, M. Sandler, and K. Cho. «Convolutional Recurrent Neural Networks for Music Classification». In: (2017), pp. 2392–2396 (cit. on p. 9).
- [16] NVIDIA. *What Are Foundation Models?* URL: <https://blogs.nvidia.com/blog/what-are-foundation-models/> (cit. on p. 9).
- [17] Meta. *AudioCraft: generating high-quality audio and music from text*. URL: <https://ai.meta.com/resources/models-and-libraries/audiocraft/> (cit. on pp. 9, 10).
- [18] J. Copet, F. Kreuk, I. Gat, T. Remez, D. Kant, G. Synnaeve, Y. Adi, and A. Défossez. «Simple and Controllable Music Generation». In: (2024) (cit. on p. 10).
- [19] Google. *MusicLM: Generating Music From Text*. URL: <https://google-research.github.io/seanet/musiclm/examples/> (cit. on p. 11).
- [20] A. Agostinelli, T. I. Denk, Z. Borsos, J. Engel, M. Verzetti, A. Caillon, Q. Huang, A. Jansen, A. Roberts, M. Tagliasacchi, M. Sharifi, N. Zeghidour, C. Frank. «MusicLM: Generating Music From Text». In: (2023) (cit. on p. 11).
- [21] OpenAI. «MuseNet». In: (2019). URL: <https://openai.com/blog/musenet> (cit. on p. 12).
- [22] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, and I. Sutskever. «Jukebox: A Generative Model for Music». In: (2020) (cit. on p. 12).

- [23] OpenAI. «Jukebox». In: (2020). URL: <https://openai.com/index/jukebox> (cit. on p. 12).
- [24] European Parliament. *UK'S APPROACH TO REGULATING THE USE OF ARTIFICIAL INTELLIGENCE*. URL: <https://www.europarl.europa.eu/topics/en/article/20230601ST093804/eu-ai-act-first-regulation-on-artificial-intelligence> (cit. on p. 24).
- [25] Mayer Brown. *AudioCraft: generating high-quality audio and music from text*. URL: <https://www.mayerbrown.com/en/insights/publications/2023/07/uks-approach-to-regulating-the-use-of-artificial-intelligence> (cit. on p. 24).
- [26] U.S. Rep. Ritchie Torres. *U.S. Rep. Ritchie Torres Introduces Federal Legislation Requiring Mandatory Disclaimer for Material Generated by Artificial Intelligence*. URL: <https://ritchietorres.house.gov/posts/u-s-rep-ritchie-torres-introduces-federal-legislation-requiring-mandatory-disclaimer-for-material-generated-by-artificial-intelligence> (cit. on p. 24).
- [27] Statista. «Generative AI - Worldwide». In: (Mar. 2024). URL: <https://www.statista.com/outlook/tmo/artificial-intelligence/generative-ai/worldwide> (cit. on p. 25).
- [28] Forbes. *The Dangers Of Inflation For Software CEOs And One Way To Fight It*. 2022. URL: <https://www.forbes.com/sites/forbesbusinesscouncil/2022/09/07/the-dangers-of-inflation-for-software-ceos-and-one-way-to-fight-it/?sh=50ea814c507d> (cit. on p. 25).
- [29] Statista. *CPI inflation rate in Europe - statistics & facts*. URL: <https://www.statista.com/statistics/685943/cpi-inflation-rate-europe/> (cit. on p. 25).
- [30] Akkio. «A Cost Breakdown of Artificial Intelligence in 2023». In: (). URL: <https://www.akkio.com/post/a-cost-breakdown-of-artificial-intelligence-in-2023> (cit. on p. 25).
- [31] Statista. *Number of social media users worldwide from 2017 to 2027*. URL: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/> (cit. on p. 25).
- [32] Oberlo. *HOW MANY PEOPLE USE SOCIAL MEDIA IN 2024?* URL: <https://www.oberlo.com/statistics/how-many-people-use-social-media> (cit. on p. 25).
- [33] Global Research and Consulting Group. *The Environmental Impact of AI*. URL: <https://insights.grcglobalgroup.com/the-environmental-impact-of-ai/> (cit. on p. 26).

## BIBLIOGRAPHY

---

- [34] AIVA. URL: <https://www.aiva.ai/> (cit. on p. 29).
- [35] Soundful. URL: <https://soundful.com/> (cit. on p. 30).
- [36] Studio by Matchtune. URL: <https://www.matchtune.studio/> (cit. on p. 30).
- [37] Dynascore. URL: <https://www.wonder.inc/dynascore/> (cit. on p. 31).
- [38] Osterwalder, A., & Pigneur, Y. «Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers». In: (2010) (cit. on p. 32).